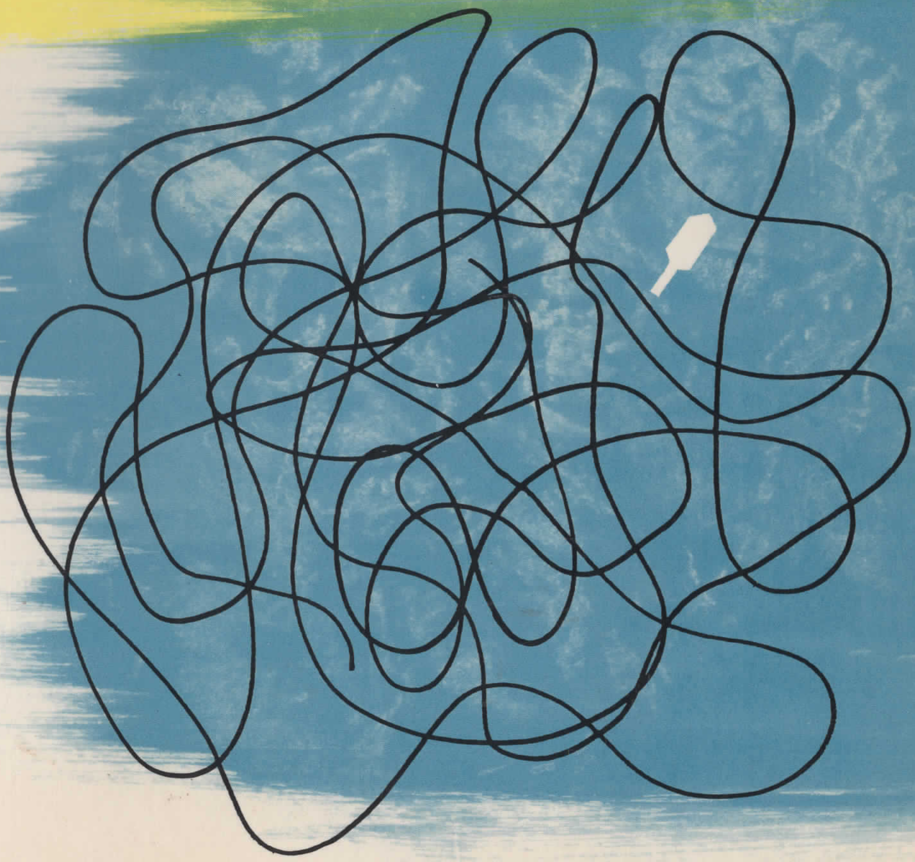
GENE ACTION

Philip E. Hartman

Sigmund R. Suskind



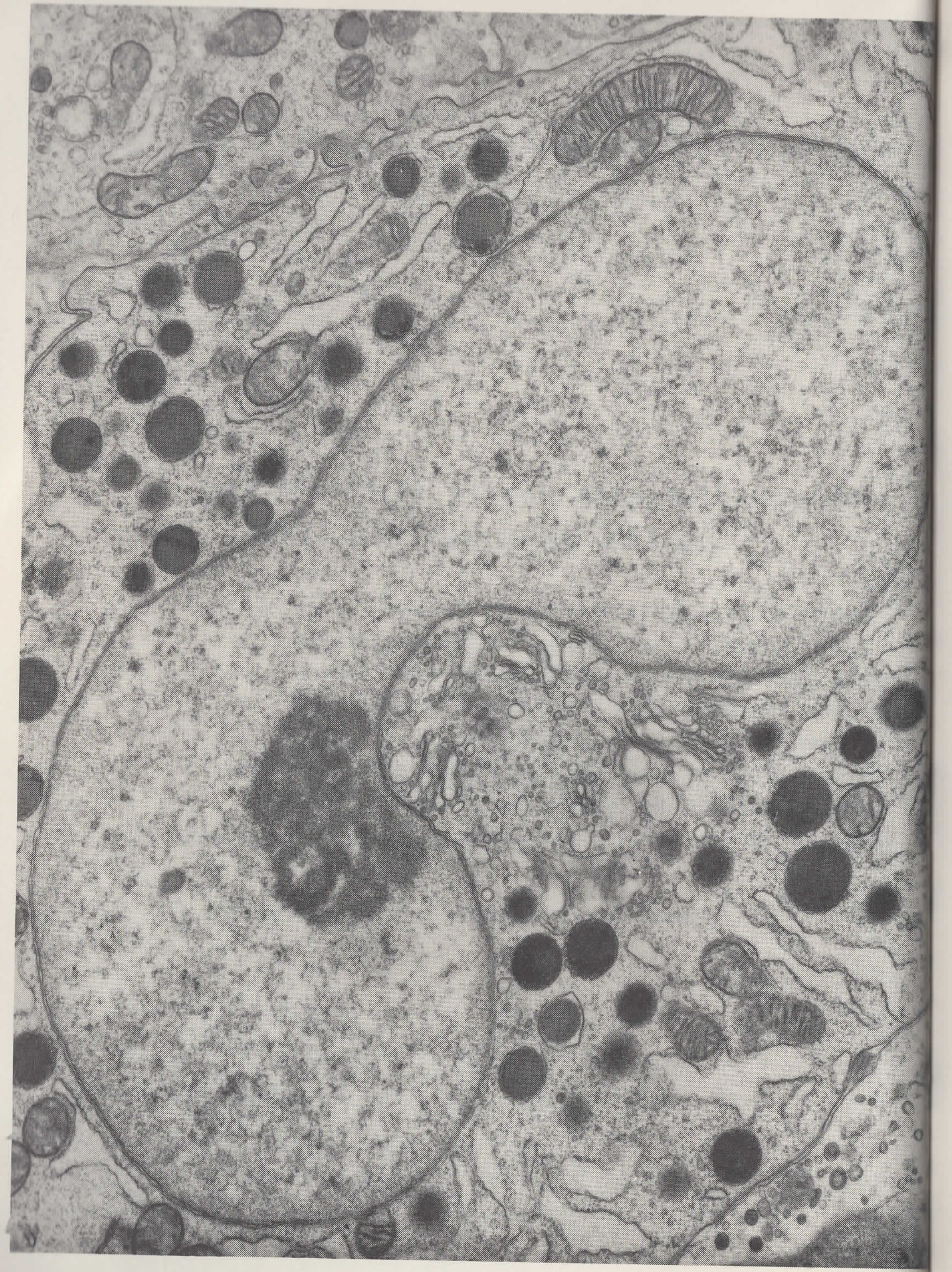
FOUNDATIONS OF MODERN Genetics SERIES

PRENTICE-HALL FOUNDATIONS OF MODERN Genetics SERIES

Sigmund R. Suskind and Philip E. Hartman, Editors

AGRICULTURAL GENETICS *James L. Brewbaker*
GENE ACTION *Philip E. Hartman and Sigmund R. Suskind*
EXTRACHROMOSOMAL INHERITANCE *John L. Jinks*
DEVELOPMENTAL GENETICS* *Clement Markert*
HUMAN GENETICS *Victor A. McKusick*
POPULATION GENETICS AND EVOLUTION *Lawrence E. Mettler*
THE MECHANICS OF INHERITANCE *Franklin W. Stahl*
CYTOGENETICS *Carl P. Swanson, Timothy Merz, and William J. Young*

* Published jointly in Prentice-Hall's *Foundations of Developmental Series*



GENE ACTION

Philip E. Hartman
Sigmund R. Suskind
The Johns Hopkins University

Prentice-Hall, Inc. Englewood Cliffs, New Jersey

FOUNDATIONS OF MODERN GENETICS SERIES

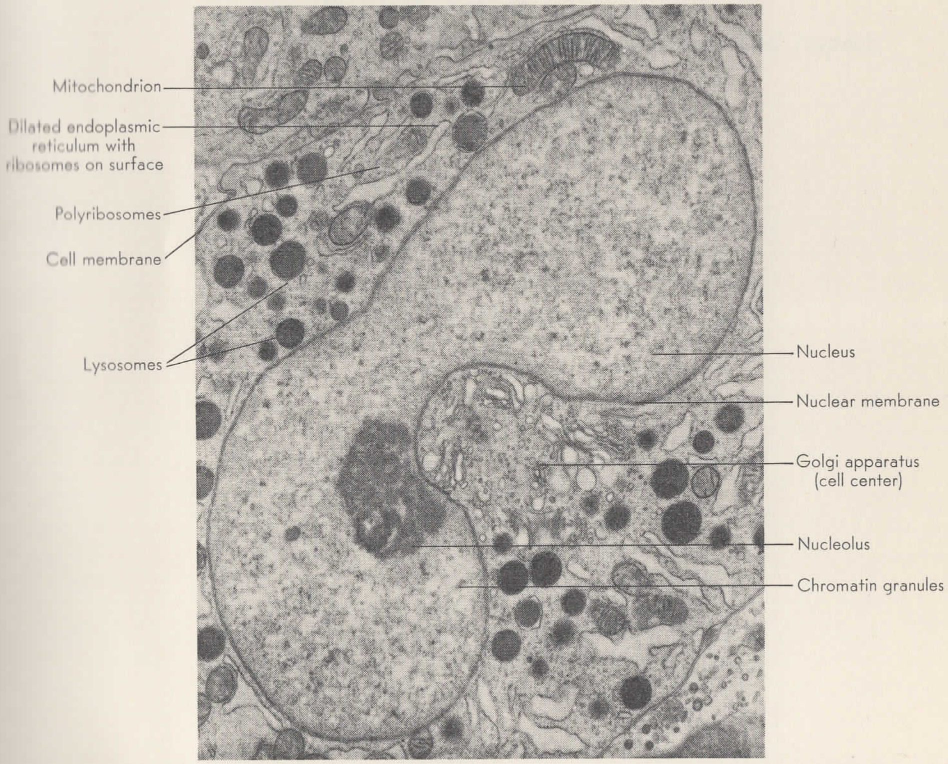
© *Copyright 1965*
by Prentice-Hall, Inc.
Englewood Cliffs, New Jersey
All rights reserved. No part of this
book may be reproduced in any form,
by mimeograph or any other means,
without permission in writing from
the publishers.

Printed in the United States of America
Library of Congress Catalog Card Number:
64-16033

Gene Action
C-34752-P
C-34753-C

Current printing (last digit):
12 11 10 9 8 7 6 5 4 3

Prentice-Hall International, Inc., *London*
Prentice-Hall of Australia, Pty., Ltd., *Sydney*
Prentice-Hall of Canada, Ltd., *Toronto*
Prentice-Hall of India (Private) Ltd., *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*


Mitochondrion
Dilated endoplasmic reticulum with ribosomes on surface
Polyribosomes
Cell membrane
Lysosomes
Nucleus
Nuclear membrane
Golgi apparatus (cell center)
Nucleolus
Chromatin granules

Frontispiece. Electron micrograph of bone marrow cell of the rat, stained with lead. Photograph courtesy of Dr. Leon P. Weiss, Department of Anatomy, The Johns Hopkins University School of Medicine. Magnification on page ii: 42,500×. Reproduced above in reduced size with identifying labels.

To David M. Bonner,
with our warm affection and respect.

Foundations of Modern Genetics

Genetic research is alive with excitement and revolutionary advances. Important to the development of science and to the evolution of social structure, genetic thought is widening its impact on many areas: immunology, protein chemistry, cellular physiology, developmental biology, medicine, agriculture, and industry.

So many partnerships and such rapidly expanding methodology demand a fresh approach to genetic training —an approach attempted in this series.

The basic principles of genetics are few and simple. We present them with enough description of accessory scientific areas to allow comprehension not only of the principles themselves but also of the types of experiments from which the concepts have evolved. Such an approach compels the reader to ask: What is the *evidence* for this concept? What are its *limitations?* What are its *applications?*

The Prentice-Hall Foundations of Modern Genetics Series presents the evidence upon which *current* genetic thought is based. It is neither a history nor a survey of all genetic knowledge. The short volumes make possible a stimulating, selective treatment of the various aspects of genetics at the intermediate level, and sectional divisions allow free choice of emphasis in differently oriented college genetics courses.

The references cited in each volume and the current research literature are the immediate sequels to this series, but the true sequel remains in the hands of the alert reader. He will find here the seed of more than one enigma, the solution of which he, himself, may help bring into man's comprehension sometime in the future.

SIGMUND R. SUSKIND
PHILIP E. HARTMAN

McCollum-Pratt Institute
The Johns Hopkins University

Genotype versus phenotype 59
Gene mutations and protein structure: Some generalities 60
    References 61
    Questions 61

Five COMPLEMENTATION 63

Getting genes together 64
An example of the *cis-trans* test for gene function 66
One gene, one function 68
One gene, one polypeptide chain 69
Intragenic complementation 70
Complementation in vitro 73
Genetic maps and complementation maps 74
Why polymers? 74
    References 75
    Questions 75

Six SECONDARY CONSEQUENCES OF GENE MUTATIONS 77

Localization of genetic blocks 77
Incomplete versus complete genetic blocks 80
Accumulations and precursors 81
Accumulations and feeding 82
Accumulations and inhibitions 83
End-product deficit 84
Feedback inhibition 84
Competition for substrate 87
Duplicate enzymes and channeled pathways 88
Consequences further removed 89
    References 90
    Questions 91

Seven MAKING THE MESSENGER 93

Watson–Crick base pairing 94
DNA-RNA hybrids 94

Two strands, one messenger 96
Messenger mutations 97
Induced messenger mutations 97
General regulation of messenger synthesis 100
Induction and repression 101
    The R *gene and the repressor* 103
    The operator 104
    The operon 105
    The unique mRNA molecule 109
    References 111
    Questions 111

Eight REGULATION ON THE RIBOSOME 112

Genes without proteins 112
Mutations and the code—some jargon 113
rII: Real genes and hypothetical polypeptide chains 114
The nucleotides between genes 115
Making sense out of missense 117
Making sense out of nonsense 118
Modulation 119
    References 121
    Questions 121

Nine THE GENETIC CODE 123

General types of codes 123
Deciphering the code 124
The coding ratio 125
Adjacent nucleotides 126
Triplet code 126
Nonoverlapping code 130
Commaless code 130
Degenerate code 131
The degeneracy may not be total 131
Quasi-universal code 132
The compositions of triplets: Amino acid incorporation studies 133
Verification of triplet composition 135
Assignments of triplets 136
    References 137
    Questions 137

Ten PERSPECTIVES AND HORIZONS 139

The eternal "why?" 140
Growth and development 141
    Gene dosage compensation 142
    Variegated position effects 143
Enzyme structure and function and the evolution of the protein molecule 146
Vitalism versus chemical reactions 146
Gene action and evolution 147
    References 147

INDEX 149

One

Base Pairing and Gene Action

Genetics currently occupies the position of keystone in the broad-based structure of the biological sciences. Cytological, developmental, physiological, biochemical, medical, ecological, and evolutionary problems increasingly are drawing from and contributing to contemporary genetic thought. This interdisciplinary stimulus, arising from the physical as well as the biological sciences, has resulted in rapid advance in the area of molecular genetics. In this book, we are concerned with the chemical nature of the genetic material and the manner in which it functions. Our approach to the problem of how genes act will be simple and direct: gene action will be equated with protein formation and function, and it will be assumed that from this all other consequences arise.

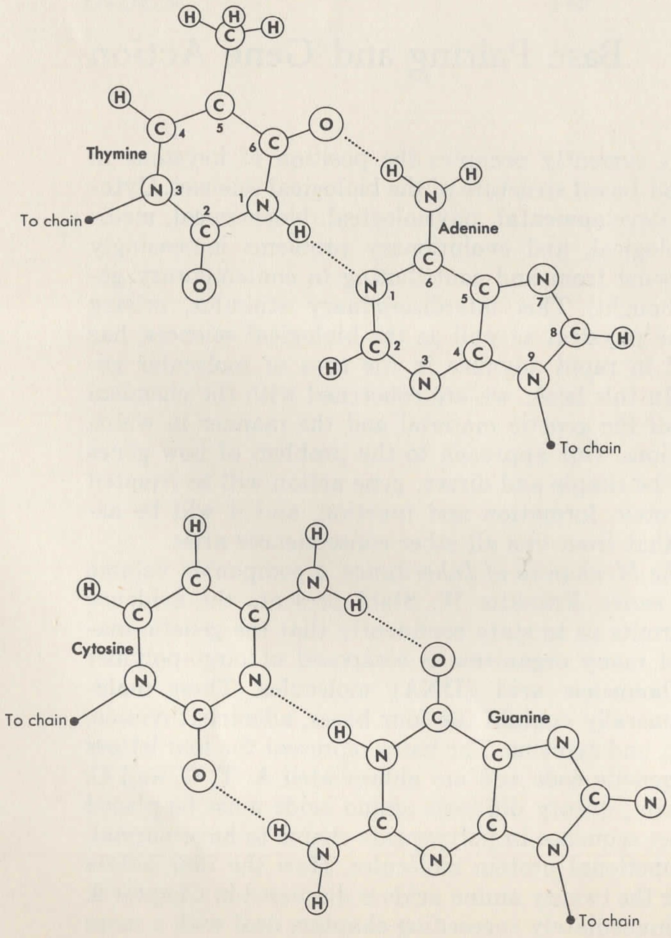
In *The Mechanics of Inheritance,* a companion volume in this series, Franklin W. Stahl presents the evidence that permits us to state confidently that the genetic material of many organisms is composed of large-polymer deoxyribonucleic acid (DNA) molecules. These molecules generally contain the four bases, adenine, thymine, guanine, and cytosine. The bases represent the four letters in the genetic code and are abbreviated A, T, G, and C. Ordinarily, twenty different amino acids must be placed in proper sequence in polypeptide chains to have normal, fully functional protein molecules. How the four letters code for the twenty amino acids is discussed in Chapter 9.

The immediately succeeding chapters deal with a more general problem: what is the protein-synthesizing machinery and how does it read and translate the nucleic acid code into an accurate positioning of amino acids in the polypeptide chain of the protein? As background for appreciation of this topic, we discuss, in Chapter **2**, protein structure and the methods by which it is elucidated.

Watson–Crick base pairing

Recall the rules of specific base pairing (Watson–Crick pairing) in double-stranded DNA: adenine pairs with thymine and guanine pairs with cytosine (Fig. 1.1). Stahl's *The Mechanics of Inheritance* presents the evidence for the existence of these hydrogen-bonded base pairs in DNA and stresses their importance in DNA replication.

Fig. 1.1. The Watson–Crick base pairs in DNA: adenine-thymine and guanine-cytosine.


Thymine
Adenine
To chain
To chain
Cytosine
Guanine
To chain
To chain

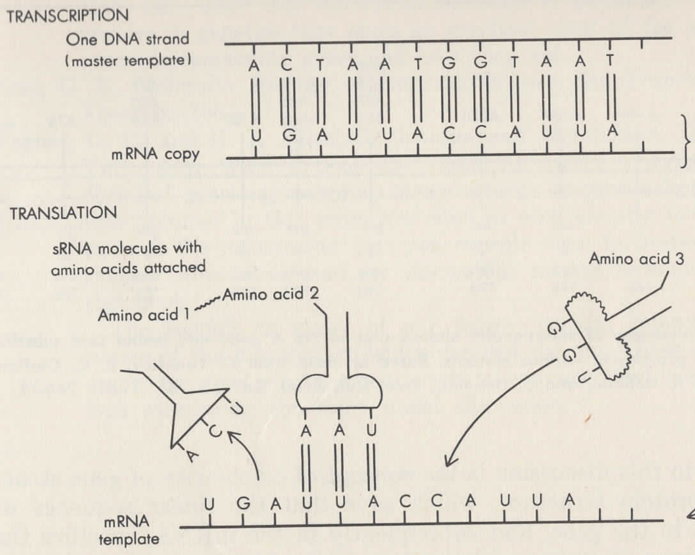

TRANSCRIPTION
One DNA strand
(master template)
A C T A A T G G T A A T
U G A U U A C C A U U A
mRNA copy
TRANSLATION
sRNA molecules with
amino acids attached
Amino acid 1
Amino acid 2
Amino acid 3
A A U
U G A U U A C C A U U A
mRNA
template

Fig. 1.2. Some principal steps in protein biosynthesis and the important roles base pairing (between adenine-uracil and guanine-cytosine) plays in the process.

Just as guanine pairs with cytosine in the two strands of the DNA double helix, so can a guanine in a DNA strand pair with a cytosine in a ribonucleic acid (RNA) strand (and vice versa). Similarly, the adenine of DNA can pair with the uracil in RNA, the thymine in DNA with the adenine in RNA. The basic hydrogen-bonding rules that apply to these DNA-RNA hybrid molecules are those that apply to the two strands of DNA itself. Two RNA strands also can form a duplex structure by similar rules.

These base-pairing rules provide a mechanism that is extremely important in the determination of the amino acid sequence of a polypeptide chain (Fig. 1.2). In protein biosynthesis, DNA (most likely only one strand of the duplex) serves as a template for a messenger RNA (mRNA) complement made in accordance with the pairing rules. The result is a DNA-RNA hybrid stage with a specific, inherited sequence of bases. The mRNA can then serve as a template upon which adaptor RNA (sRNA) molecules are oriented, in a sequence that also is determined by the Watson–Crick pairing rules. As a result of this step, the individual amino acids can be ordered in a specific sequence (or primary structure), for each sRNA molecule carries a specific amino acid. Finally, the amino acids are linked by peptide bonds to form a polypeptide chain. The amino acid sequence of this chain, then, has been dictated by the original DNA code, transcribed through messenger RNA, and translated by adaptor RNA through the mechanism provided by the base-pairing rules.

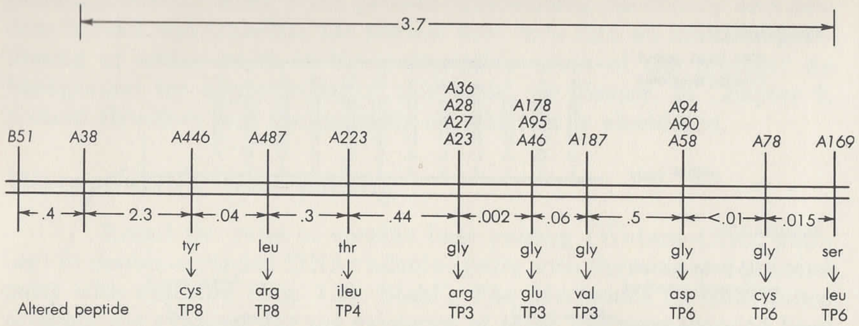


Fig. 1.3. The colinearity of mutationally altered sites in the *A* gene and amino acid substitutions in the A proteins of various mutants. Based on data from C. Yanofsky, B. C. Carlton, J. R. Guest, D. R. Helinski, and U. Henning, *Proc. Nat. Acad. Sci. U.S.*, *51* (1964), 266-72.

Implicit in this discussion is the concept of colinearity of gene structure and protein structure, which says that the linear sequence of nucleotides in the gene, and subsequently in the mRNA, specifies the linear sequence of amino acids in the protein. Until recently, this concept was unsupported by direct genetic and biochemical evidence. However, through the elegant studies of C. Yanofsky and his coworkers, strong evidence is now available for the *colinearity* of amino acids in the bacterial enzyme, tryptophan synthetase, and the gene that controls its structure. This has been accomplished, as shown in Fig. 1.3, by demonstrating that recombination distances between positions in the gene parallel the distances between affected amino acid residues in the protein. You might think about the importance and the significance of this observation as you read the chapters that follow.

In the next two chapters, we shall consider some of the basic chemistry of protein structure as well as certain aspects of protein biosynthesis you will need in later discussions.

References, including extensive introductions to microbial and biochemical genetics

Adams, M. H., *Bacteriophages*. New York: Interscience Publishers, Inc., 1959.

Burdette, W. J., ed., *Methodology in Basic Genetics*. San Francisco: Holden-Day, Inc., 1963.

Fincham, J. R. S., and P. R. Day, *Fungal Genetics*. Philadelphia: F. A. Davis, Co., 1963.

Hayes, W., *Genetics of Bacteria and Their Viruses*. New York: John Wiley & Sons, Inc., 1964.

Jacob, F., and E. L. Wollman, *Sexuality and Genetics of Bacteria*. New York: Academic Press, Inc., 1961.

Peters, J. A., *Classic Papers in Genetics*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1959. An annotated collection of important original literature in genetics that gives an excellent view of the evolution of concepts concerning genes and how they act.

Stent, G. S., *Molecular Biology of Bacterial Viruses*. San Francisco: W. H. Freeman, 1963.

Wagner, R. P., and H. K. Mitchell, *Genetics and Metabolism*, 2nd ed. New York: John Wiley & Sons, Inc., 1964. An excellent textbook on biochemical genetics, including many aspects of gene action.

NOTE: Other volumes in this series are cited in each chapter to which they pertain. We recommend that you consult them for important background information and for discussions related to many topics in this book.

The legends to many of our figures contain references to the source of the material explicitly detailed therein. These sources generally are good starting points where you may find more information concerning the topic under discussion.

Two

Protein Structure

From both a structural and a functional standpoint, proteins are a most important chemical constituent of cells. Enzymes, the proteins with catalytic activities, are largely responsible for the phenotype or properties of a cell in a particular environment. The genotype of a cell, its total genetic material, determines the *potential* types and relative amounts of protein the cell can produce. Thus, the environment influences the *expression* of the genetic potentialities of the cell, while the genotype sets limits on these potentialities.

In considering the types of proteins produced by a cell, we must take into account not only the particular catalytic activities of the proteins, but also other important features and properties of their structure that profoundly affect their role in cellular physiology. Proteins are immensely versatile macromolecules. Built into their structure is information that instructs them in "what to do" (catalytic activity), "where to go" in the cell (intracellular organization), and "when and how to perform" (control of function through interaction with other proteins, small activators, or inhibitors).

A single mutation in the genetic material may lead to a multitude of diverse effects on the phenotype of an organism. When these phenotypic changes have been traced back through the complex web of metabolism, they have been found to result from a simple but critical change in the structure of a single protein. Thus, the ultimate mutant phenotype is a secondary consequence of the ramifications and interactions occurring in cellular metabolism that arise from this initial alteration.

At the present time, most gene action can be attributed to a mechanism by which individual genes dictate the structures of individual kinds of proteins. To understand this theory, one must have some knowledge about the structure of the genetic material and the structure of proteins and about the biochemical mechanisms involved in protein biosynthesis. This chapter summarizes some aspects of protein structure essential to the later discussions of protein biosynthesis, mutational alterations in protein structure, and the guiding role genetic material plays in the biosynthesis of particular proteins.

Primary structure

The *primary* structure of a protein refers to the sequence of small, nitrogen-containing compounds called amino acids that are the building blocks of the protein molecule. In the general structure for an ionized amino acid shown in Fig. 2.1, R stands for side-chains that are different for each amino acid. R can be as simple as a hydrogen atom (H) or a methyl group (CH_3), or it can be a more complex structure. The first carbon is part of the acidic carboxyl group (COO^-). The second carbon (printed in bold type), to which is attached a basic ammonium group (NH_3^+), is called the alpha carbon. Amino acids are amphoteric electrolytes: they can act either as acids or as bases. The amino acid in Fig. 2.1 is in the zwitterion, or dipolar, form. In acid solutions, the ionization of the carboxyl group is suppressed, and the molecule possesses a positive charge; in basic solutions, the molecule loses a proton from the amino group and thus carries a negative charge. The net charge of the protein, determined by the charge of the individual R groups, is responsible for the electrolytic properties of the molecule, so that proteins, as well as their constituent amino acids, are amphoteric molecules.

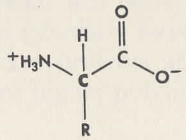
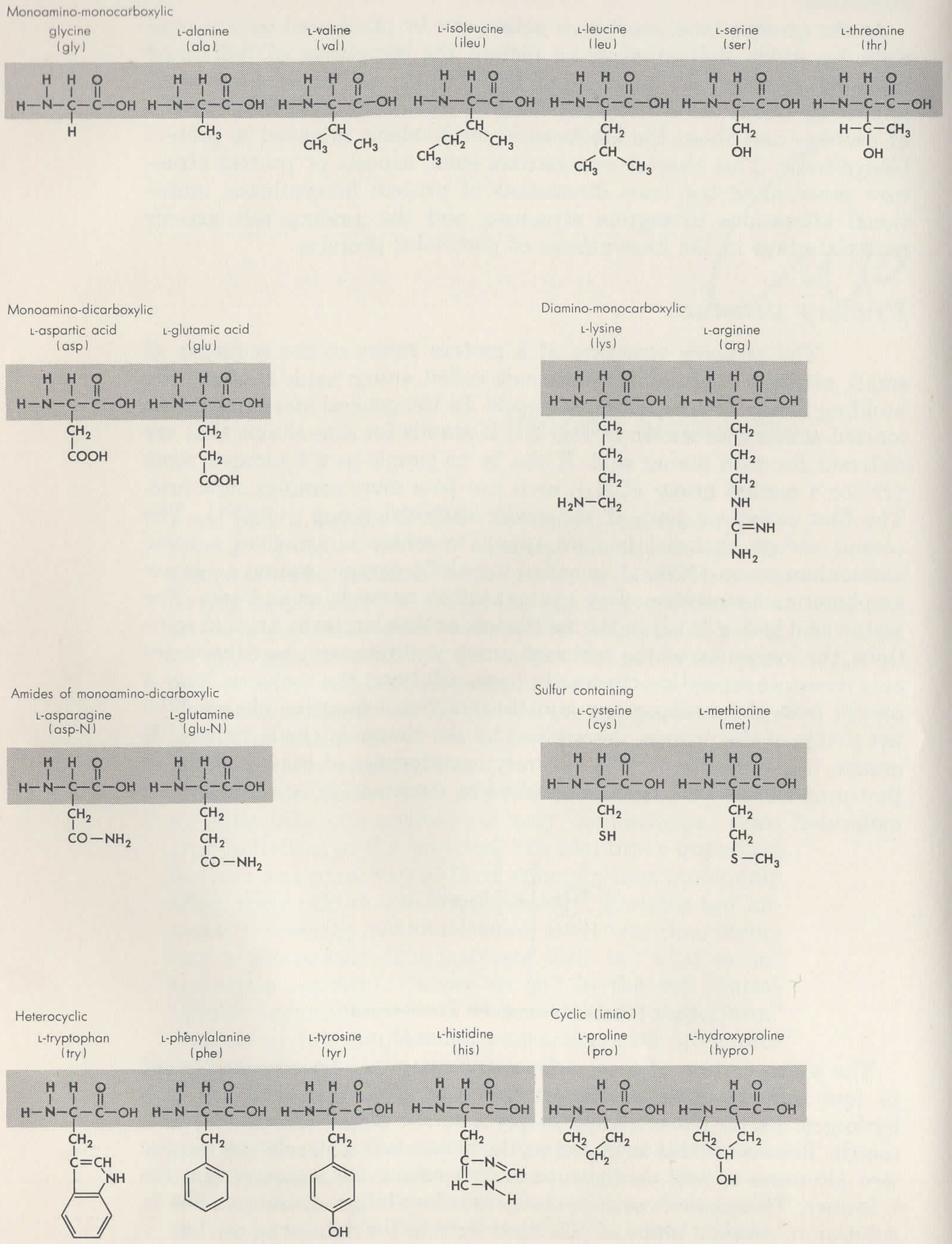


Fig. 2.1. General formula for an amino acid.

The alpha carbon of most amino acids is joined by covalent bonds to four different groups. The exception is glycine, in which R is a hydrogen atom. Thus, except in glycine, the alpha carbon is asymmetric. Because of this asymmetry, the amino acid molecule can assume two alternate optical configurations or isomers, the D isomer and the L isomer. This optical activity refers to the ability of amino acids in solution to rotate a plane of polarized light to the right or to the left.

Fig. 2.2. Amino acids found in proteins.

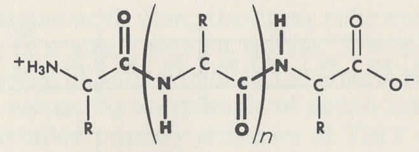

Monoamino-monocarboxylic
glycine (gly)
L-alanine (ala)
L-valine (val)
L-isoleucine (ileu)
L-leucine (leu)
L-serine (ser)
L-threonine (thr)
Monoamino-dicarboxylic
L-aspartic acid (asp)
L-glutamic acid (glu)
Diamino-monocarboxylic
L-lysine (lys)
L-arginine (arg)
Amides of monoamino-dicarboxylic
L-asparagine (asp-N)
L-glutamine (glu-N)
Sulfur containing
L-cysteine (cys)
L-methionine (met)
Heterocyclic
L-tryptophan (try)
L-phenylalanine (phe)
L-tyrosine (tyr)
L-histidine (his)
Cyclic (imino)
L-proline (pro)
L-hydroxyproline (hypro)

Although many different amino acids have been isolated from various plants and animals, only twenty of the L isomers are found in most proteins. Some polypeptides, mainly constituents of microbial cell walls, do contain D amino acids. The nearly universal use of only twenty amino acids for the synthesis of widely different proteins is one of nature's enigmatic rules. We can only argue, teleologically, that during the course of evolution other amino acids may have been tried, tested, and found to present either no advantage or a disadvantage to organisms.

Formulas for these twenty amino acids, as well as for the less frequently occurring hydroxyproline, are depicted in Fig. 2.2. The abbreviations given in the figure are standard nomenclature, agreed upon by chemists and increasingly utilized by geneticists.

Proteins vary widely in amino acid content. Different types of proteins within an organism may have very different amounts of a particular amino acid. Some amino acids are numerous in one protein, but rare in others. Tryptophan, for example, is entirely absent in certain proteins. Usually, however, a single protein lacks no more than one or two of the twenty amino acids.

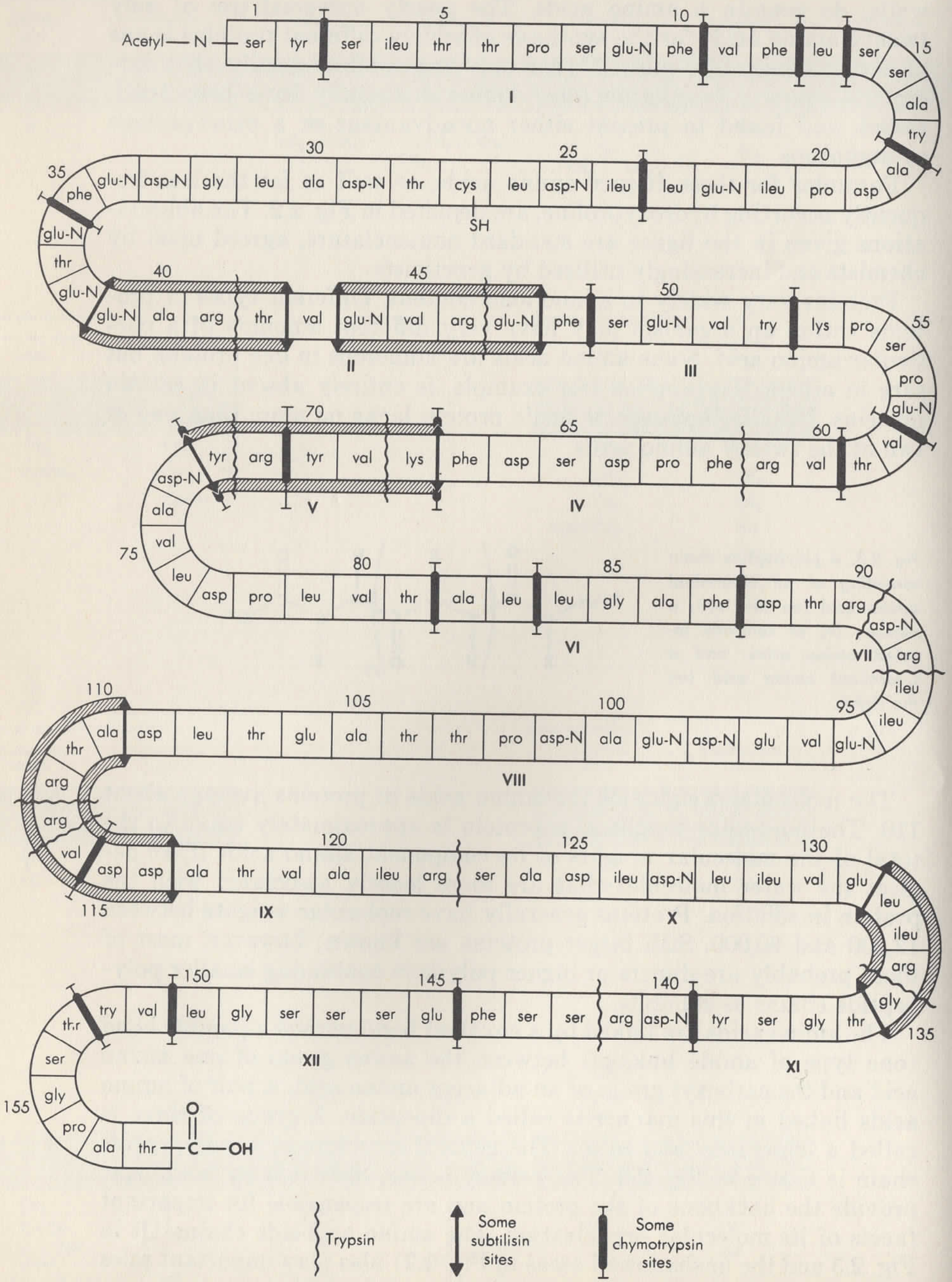
Fig. 2.3. A polypeptide chain consisting of an N-terminal amino acid (on the left), a number (n) of centrally located amino acids, and a C-terminal amino acid (on the right).



The molecular weights of the amino acids in proteins average about 110. The molecular weight of a protein is approximately equal to the total of the molecular weights of its component amino acids if we neglect the water molecules that are more loosely associated with the protein in solution. Proteins generally have molecular weights between 12,000 and 90,000. Still larger proteins are known; however, most of these probably are dimers or higher polymers containing smaller polypeptide chains as subunits.

The amino acids are linked by a covalent bond termed a peptide bond (one type of amide linkage) between the amino group of one amino acid and the carboxyl group of an adjacent amino acid. A pair of amino acids linked in this manner is called a dipeptide. A group of three is called a tripeptide, and so on. The general structure of a *polypeptide* chain is shown in Fig. 2.3. The *peptide bonds,* indicated by bold type, provide the backbone of the protein and are responsible for important facets of its molecular organization. The amino acid side chains (R in Fig. 2.3 and the unshadowed areas in Fig. 2.2) also play important roles in protein structure and function.

Fig. 2.4. Primary structure of tobacco mosaic virus (TMV) protein. Some of the subtilisin peptides (shaded positions in the figure) overlap, or bridge, the ends of some tryptic peptides and have assisted in ordering the sequence of these tryptic peptides.


Acetyl—N—
1 ser tyr ser ileu 5 thr thr pro ser glu-N 10 phe val phe leu ser 15
I
ser ala try ala asp pro 20 ileu glu-N leu ileu 25 asp-N leu cys thr asp-N 30 ala leu gly asp-N glu-N 35 phe
SH
glu-N thr glu-N
40 glu-N ala arg thr val 45 glu-N val arg glu-N phe 50 ser glu-N val try lys pro 55
II
III
ser pro glu-N val thr 60 val arg phe pro asp 65 ser asp phe lys val 70 tyr arg tyr
IV
V
asp-N ala 75 val leu asp pro 80 leu val thr ala leu 85 leu gly ala phe asp thr 90 arg
VI
asp-N arg ileu ileu 95 glu-N val glu asp-N glu-N 100 ala asp-N pro thr thr 105 ala glu thr leu asp 110 ala
VII
VIII
thr arg arg val 115 asp asp ala thr val 120 ala ileu arg ser ala 125 asp ileu asp-N leu ileu 130 val glu
IX
X
leu ileu arg gly 135 thr gly ser tyr 140 asp-N arg ser ser phe 145 glu ser ser ser gly 150 leu val try
XI
XII
thr ser gly 155 pro ala thr—C—OH
O
Trypsin
Some subtilisin sites
Some chymotrypsin sites

The terminal amino acid with the free amino group in the polypeptide chain is called the N-terminal amino acid. The terminal amino acid with the free carboxyl group is called the C-terminal amino acid. Each amino acid in the chain is termed a residue. In numbering the amino acids in an extended polypeptide chain, convention dictates that the N-terminal residue be number 1 and that the C-terminal residue be the final number in the sequence. When the sequence of amino acids in a peptide is described, the C-terminal residue is mentioned last. For example, a tripeptide with an N-terminal glycine linked to an alanine that is linked to a C-terminal histidine is referred to as glycyl-L-alanyl-L-histidine and is abbreviated as gly.ala.his. If the sequence of amino acids in such a tripeptide is not known, the abbreviation would be (gly. ala.his), the parentheses denoting that only the composition of the tripeptide is known.

A single protein species produced by genetically identical cells or by genetically identical virus particles has a definite composition. For example, each of the protein molecules contained in a normal tobacco mosaic virus (TMV) particle contains 158 amino acid residues, each amino acid being represented a certain number of times. When different batches of normal protein of TMV are analyzed for their amino acid composition, a characteristic number of each type of amino acid residue is found. The sequence of amino acids along the linear polypeptide chain is unique and is constant in a single protein species. The stability of both composition and sequence of amino acids in the protein remains as long as no genetic change occurs. As an example of amino acid composition and sequence, the detailed primary structure of TMV protein is shown in Fig. 2.4.

Determining the sequence of amino acids

Highly refined chemical and enzymatic procedures, stemming from those introduced by Nobel prize winner Frederick Sanger in the 1940's, have been used to elucidate the primary structure of proteins. Understanding the details of gene action involves a knowledge of the relative positions of amino acids in a specific protein. We shall examine some of the basic procedures for determining the amino acid and peptide sequence of a protein.

The long polypeptide chain cannot be analyzed directly, for the chemical methods of splitting off amino acids one at a time (discussed below) lose some resolution at each step. The treatments break some of the internal peptide bonds, and the attack on all the molecules of a protein population is not entirely synchronous or complete. Therefore, the polypeptide first must be cut into short peptide segments that are susceptive to stepwise treatments.

Fragmentation of protein molecules. To fragment the polypeptide chains of a population of identical protein molecules, the breakages must be obtained at the same specific positions in each of the chains. Only in this manner can a yield of fragments of homogeneous composition be produced. Several chemicals can be used to accomplish this. For

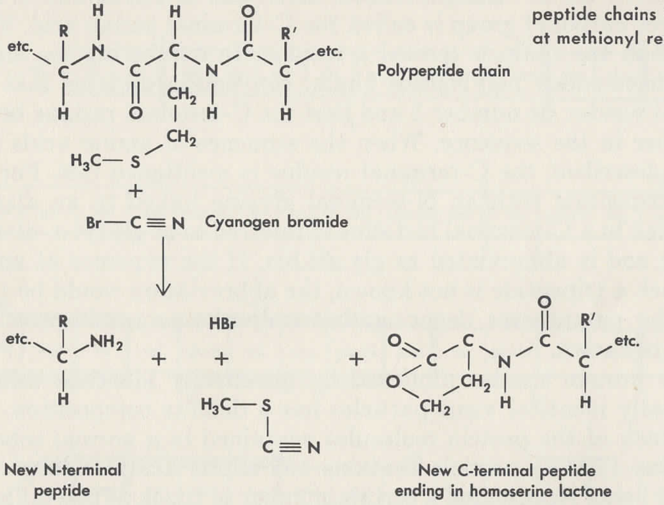


Fig. 2.5. Chemical treatment that specifically breaks polypeptide chains into fragments at methionyl residues.

example, cyanogen bromide in slightly acidic solution cleaves the peptide bond between methionine and the next amino acid, yielding two peptide fragments, one with a new N-terminal amino acid and one with a C-terminal homoserine (Fig. 2.5). Another useful chemical is the oxidizing agent N-bromosuccinimide, which attacks tryptophan residues. In TMV protein, the reaction occurs at positions 17, 52, and 152 (Fig. 2.4), yielding four peptide fragments.

The proteolytic enzyme trypsin specifically splits polypeptide chains at the carboxyl side of arginine and lysine residues (shown by wavy lines in Fig. 2.4) unless proline is a succeeding residue (e.g., residues 53, 54). The peptides so released (tryptic peptides) are indicated by Roman numerals, numbering from the N-terminal to the C-terminal end of the original polypeptide chain. The production of a mixture of the two types of tryptic peptides can be avoided by pretreating the protein with carbon disulfide. This compound chemically protects the lysine residues from the action of trypsin, and thus tryptic hydrolysis occurs only at the arginine sites.

Another widely used enzyme is chymotrypsin, which most effectively hydrolyzes peptide bonds adjacent to the aromatic amino acids tryptophan, phenylalanine, and tyrosine; it hydrolyzes other peptide bonds more slowly. Some known sites of hydrolysis by chymotrypsin are indicated with solid bars in Fig. 2.4. The hydrolysis is blocked if a proline residue is next to the aromatic amino acid (for example, the peptide bond between residues 62 and 63 in Fig. 2.4). Other proteolytic enzymes also may be used—pepsin or subtilisin, for example. Figure 2.6 lists the peptides of TMV protein obtained by hydrolyses with these four enzymes.

Frequently, the proteolytic enzymes are used sequentially. Thus, trypsin can be used first to produce tryptic peptides, and these then can be exposed to chymotrypsin for reduction to smaller peptide fragments.

Isolation and characterization of peptides. As can be seen in Fig. 2.4, treatment of a protein can produce a mixture of peptides of different amino acid composition. Therefore, before the individual peptides can be analyzed, they must be separated.

The methods for separation take advantage of the different sizes and solubilities of the various peptide fragments, as well as of their different amounts of charged amino acid residues. Separation and purification can be achieved by passing the protein digest through a glass column filled with an ion-exchange resin. Such a resin has a higher affinity for some charged groups than it has for others; thus it retards the passage through the column of some peptides to a greater extent than it does others. Peptides also may be separated on the basis of their differential solubilities in certain solvents by a process called countercurrent distribution. A mixture of peptides also can be separated by the "fingerprint" technique, which selectively segregates different peptides and at the same time characterizes them on the basis of their physical and chemical properties. The technique consists of three procedures. Figure 2.7 shows a fingerprint of peptides obtained from a digest of normal tryptophan synthetase (see Chapter 4). A small drop of the protein hydrolysate was placed at a spot (labeled *origin* in the figure) on one edge of a sheet of filter paper. A suitable solvent was allowed to flow through this spot and across the paper, moving different peptides different distances, depending on the solubilities of the peptides in the solvent used. This first step is known as chromatography. The paper was then turned 90 degrees and placed in a second solvent, and a high-voltage electric current was run through the solution. Because the peptides differed in net charge, they migrated at dif-

Fig. 2.6. Some peptides obtained from TMV protein after digestion with various proteolytic enzymes. The numbers refer to the amino acids in the TMV protein, shown in Fig. 2.4. The precise amino acid content of each class of peptide fragment has been determined, as has, in most cases, the sequence of amino acids in the fragment.

	Trypsin	Subtilisin	Chymotrypsin	Pepsin
I.	1-41	39-42	46-48	35-42
II.	42-46	44-47	53-57 53-59	36-42
III.	47-61	68-72	53-67	40-42
IV.	62-68	110-116	88-114	43-47
V.	69-71	132-135	140-144	45-47
VI.	72-90			43-50
VII.	91-92			51-59
VIII.	93-112			60-66
IX.	114-122			68-75
X.	123-134			109-120
XI.	135-141			121-128
XII.	142-158			133-138
				133-143

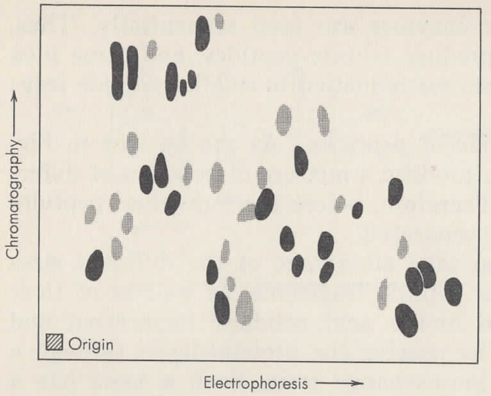


Fig. 2.7. Fingerprint of peptide fragments obtained after tryptic and chymotryptic digestion of wild-type *Escherichia coli* tryptophan synthetase A protein. The intensity of the staining of a spot is influenced by the quantity and composition of the peptide. See text for details. Drawn from D. R. Helinski and C. Yanofsky, *Proc. Nat. Acad. Sci. U.S.,* 48 (1962), 173.

ferent rates. This step is called electrophoresis. The combination of the two procedures conveniently separates the different peptides by localizing them at different spots on the paper.

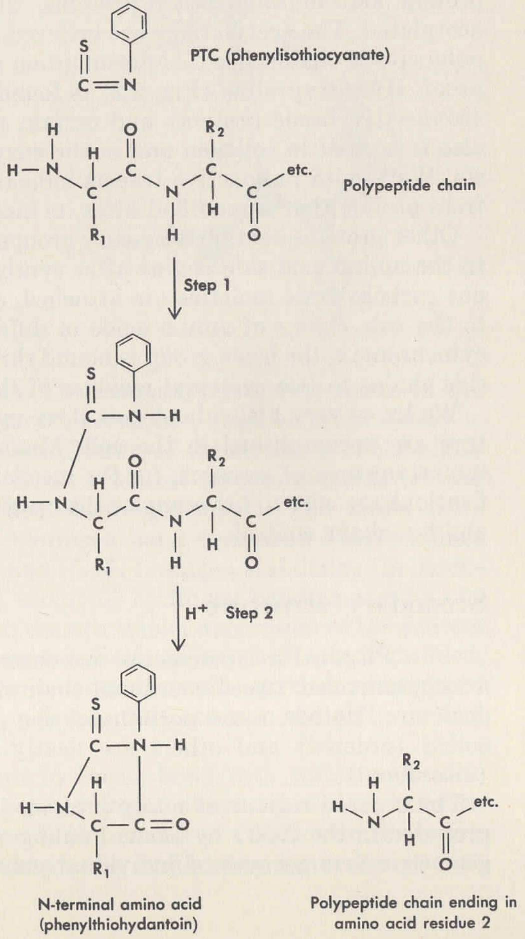
The paper was then sprayed with a dye, ninhydrin, which revealed the distribution of the peptides. This dye, which forms blue complexes with alpha amino acids, reacts with the peptides. To obtain still more information about the amino acid composition of the peptides, other color reagents can be used, each of which reacts with and thus reveals only those peptides that contain histidine, arginine, or tyrosine residues.

Analysis of peptides. Now, the peptides having been isolated and characterized, they can be analyzed for their individual content and sequence of amino acids. Enzymes called exopeptidases are useful in this analysis. An exopeptidase splits off amino acids one at a time from one of the ends of the peptide. The sequence and nature of the amino acids composing the peptide influence the enzymatic release of free amino acids. For example, the specificity of aminopeptidase disallows release of a substituted N-terminal amino acid. No N-terminal amino acid was found after treatment of TMV protein; then it was discovered that acetyl serine, a substituted amino acid, occupies the first position in this protein (Fig. 2.4). Some other proteins also have acetyl serine or another acetylated amino acid in the N-terminal position. As a further example of how structure affects enzymatic digestion, we may observe the unique bonding in the polypeptide chain at proline residues (see Fig. 2.2). When proline becomes the N-terminal residue during digestion, the enzyme leucine aminopeptidase is unable to hydrolyze further. This configurational specificity also affects hydrolysis by carboxypeptidase A. For example, the proline at position 156 in TMV protein (Fig. 2.4) limits hydrolysis by carboxypeptidase A to the terminal threonine residue.

A chemical method superior to the use of aminopeptidase for the stepwise degradation from the N-terminal end is phenylisothiocyanate, depicted in Fig. 2.8. The N-terminal amino acid reacts with phenylisothiocyanate in step 1. Treatment with acid splits off the derivative, phenylthiohydantoin amino acid. This can be separated, and the remaining polypeptide chain then can be recycled for the determination of the second amino acid in the sequence (now the N-terminal amino acid).

Putting the pieces together. To determine the order of the amino acids in the entire chain, the order of the peptides in the chain must be established. The task is lightened because the peptides overlap in their amino acid sequence. A simple example of the usefulness of overlapping sequence can be visualized by attempting to spell a ten-letter word correctly, given only sequences of three, four, or five letters at a time. If there is some overlap of letters within the groups, it becomes much simpler to put the pieces together to form the intact word. The overlapping in TMV peptides can be seen in Figs. 2.4 and 2.6. By comparing the sequences of different peptides, the linear order of the peptides in

Fig. 2.8. Edman PTC method for the determination of N-terminal amino acid residues in peptides.


PTC (phenylisothiocyanate)
Polypeptide chain
Step 1
H+ Step 2
N-terminal amino acid (phenylthiohydantoin)
Polypeptide chain ending in amino acid residue 2

TMV protein was obtained. We shall take for an example only the tryptic peptides IV, V, and VI and the associated subtilisin peptide. The latter overlaps the ends of tryptic peptides IV and V. Note that the only lysine present in the protein that can terminate a tryptic peptide is that of tryptic peptide IV. This same lysine residue constitutes the C-terminal residue of one of the subtilisin peptides that also contains a valine, a tyrosine, an arginine, and another, terminal tyrosine. This overlapping indicates that the small tryptic peptide containing valine, tyrosine, and arginine follows tryptic peptide IV and must be V. It also indicates that the only tryptic peptide with the C-terminal tyrosine must follow tryptic peptide V and be peptide VI.

Comparison of the compositions and sequences of different peptides has permitted the construction of the complete unique linear sequence of amino acids in TMV protein and a small number of other proteins.

Modifications in primary structure

Some proteins contain rare amino acids. For example, in TMV protein and in some other proteins, the N-terminal amino acid is acetylated. The acetylations are believed to occur after synthesis of the polypeptide chains, but this presumption awaits definitive experimental proof. Hydroxyproline (Fig. 2.2) is found in large amounts in collagen (connective-tissue protein) and certain plant proteins; hydroxylysine also is present in collagen and in the enzymes trypsin and chymotrypsin. Work with radioactive tracers indicates that hydroxyproline arises from proline that is modified after its incorporation into collagen.

Other proteins contain accessory groups that are believed to be added to the amino acid side chains after synthesis of the polypeptide. Various carbohydrate moieties are attached, often through covalent bonds, to the side chains of amino acids in different proteins. In the enzyme cytochrome *c*, the heme group is bound through union of two of its vinyl side chains to two cysteinyl residues of the polypeptide chain.

We know very little about how these modifications in primary structure are accomplished in the cell. Much remains to be done in this important area of research, for the mechanisms underlying these modifications are sure to have important bearing on concepts of gene action and metabolic control.

Secondary structure

Physical measurements demonstrate that a protein is neither a long, extended, two-dimensional chain nor simply a randomly coiled structure. Rather, some portions of the polypeptide chain are neatly coiled (ordered) and others are neatly folded but remain uncoiled (disordered).

The ordered regions of many proteins have the secondary structure proposed in the 1950's by Linus Pauling and his coworkers—a definite geometric arrangement of individual polypeptide chains stabilized by a variety of noncovalent bonds. The secondary structure found in the ordered regions of most proteins (the compact, or globular, proteins) is a helical coil called an alpha helix. Details of the structure of the alpha helix are shown in Fig. 2.9. Such a helical structure meets certain essential, restrictive physicochemical requirements. The alpha helix allows maximum intrahelical hydrogen-bond formation (broken lines in Fig. 2.9) between the C=O and H—N linkages, stabilizing the structure in the helical form. The extent of hydrogen bonding can be observed by noting the changes in the ultraviolet absorption of the protein solution when the protein is in a helical, as opposed to an uncoiled, form. In addition, optical rotation measurements give an estimate of the proportion of the protein molecule that exists in the alpha-helical configuration in solution.

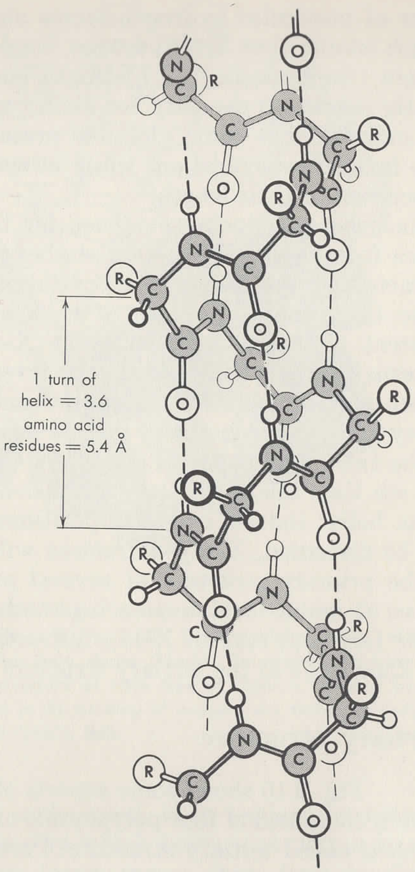


Fig. 2.9. Portion of a polypeptide chain coiled in the alpha-helical configuration, drawn to indicate the three-dimensional structure of the helix. Hydrogen bonds are indicated by broken lines, hydrogen atoms by small open circles. C, N, and O represent carbon, nitrogen, and oxygen atoms, respectively; R stands for the side chains of the amino acids composing the peptide chain. Redrawn from Linus Pauling, *The Nature of the Chemical Bond*, 3rd ed. (Cornell, 1960).

In aqueous solution, proteins hydrogen-bond with water molecules, and, in the process, an exchange of hydrogen atoms takes place between the water and the protein. Using deuterium water (deuterium is a heavy hydrogen isotope), one can measure the availability for exchange with water of particular hydrogen atoms along the polypeptide chain. Hydrogen atoms that are hydrogen bonded within the structure of the protein (for example, the hydrogen bonds of the alpha helix) have a greatly decreased capacity for deuterium exchange. All of these methods indicate that some globular proteins may be almost completely in a helical configuration, while others are ordered over as little as 20 per cent of their length.

The most convincing evidence for the presence of the alpha helix comes from X-ray diffraction studies on protein crystals. X rays are diffracted by the planes of a crystalline structure of a substance. Because the atomic structure of each crystalline substance is unique, different molecules give different X-ray diffraction patterns. Such patterns from protein crystals can reveal their molecular arrangement. Nobel prize winners J. C. Kendrew and M. F. Perutz have developed X-ray diffraction methods that allow resolution of protein structure at the two-angstrom level. From the dimensions given in Fig. 2.9, one can see that this resolution is sufficient to probe the structure of the alpha helix; indeed, even the positions of the individual amino acids can be discerned. X-ray diffraction will ultimately allow confirmation of the primary structure of several proteins previously analyzed by chemical amino acid–sequencing methods. Of very great importance is the fact that current X-ray diffraction methods are contributing to our knowledge of the tertiary structure of proteins.

Tertiary structure

Fig. 2.10 shows some aspects of the structure of hemoglobin, a protein composed of four polypeptide chains folded into a configuration which is called tertiary structure. Perhaps more strikingly, Fig. 2.10b pictures the four polypeptide chains of the protein molecule, each with its own primary, secondary and tertiary structure, arranged in a three-dimensional configuration termed quaternary structure (see p. 21).

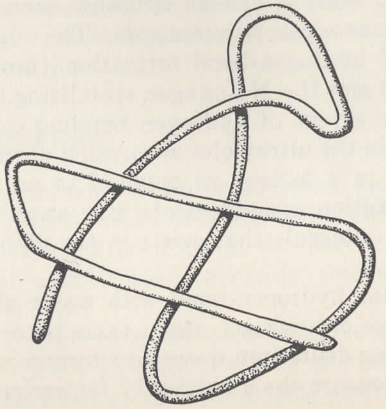


Fig. 2.10a. One chain of the hemoglobin molecule, showing its folded structure. Hemoglobin is a blood protein that binds oxygen reversibly and thus carries a supply of oxygen to the tissues. Redrawn from M. F. Perutz, *Science, 140* (1963), 863.

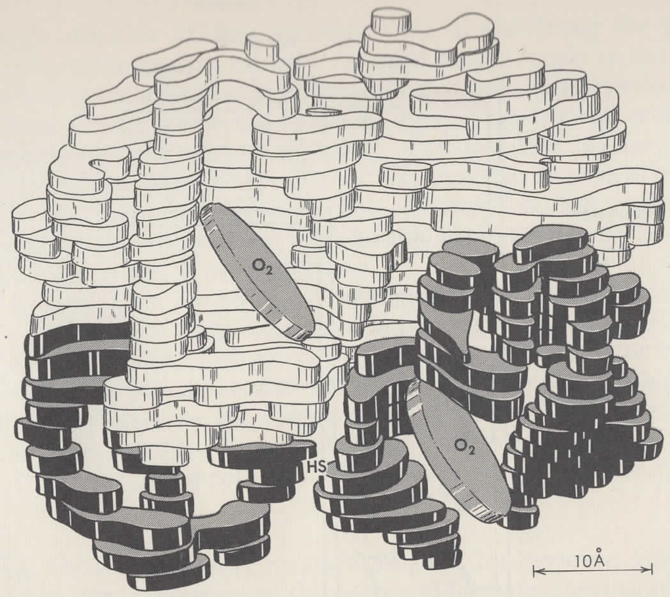


Fig. 2.10b. Interaction of four polypeptide chains to constitute the hemoglobin molecule. Each of the chains is folded approximately as in Fig. 2.10a. The two alpha chains (white) are alike in primary structure, as are the two beta chains (black). The alpha and beta chains differ slightly from each other, for the structure of each type of chain is dictated by a different gene. The heme groups, associated in the binding of oxygen, are shown as gray disks. Redrawn from M. F. Perutz, *Science, 140* (1963), 863.

We are just beginning to understand how tertiary structure is achieved and the types of forces by which tertiary structure is maintained. Our knowledge of these forces comes, first, from studies involving the disruption (denaturation) of the complex configuration of the protein and its restoration (renaturation) to a native configuration. Some of the forces that cause and maintain the folded configuration are results of various types of interactions of amino acid side chains. Second, in three-dimensional models constructed from X-ray diffraction patterns, one can determine which amino acid side chains in one segment of the polypeptide chain are in juxtaposition with amino acid side chains of another portion of the polypeptide chain.

As you will recall, hydrogen bonding between C=O and H—N groups within the helical structure of the polypeptide is a major stabilizing force in maintaining this secondary structure. Certain difficulties arise, however, in the proximity of proline residues within the polypeptide chain. The cyclical nature of proline and the resulting imino peptide bonds are unique among the twenty amino acids most widely found in proteins (see Fig. 2.2). No hydrogen bonds involving imino peptides occur. The bond angles necessary for rotation of the molecule into the helix are restricted at proline residues: the cyclical nature of

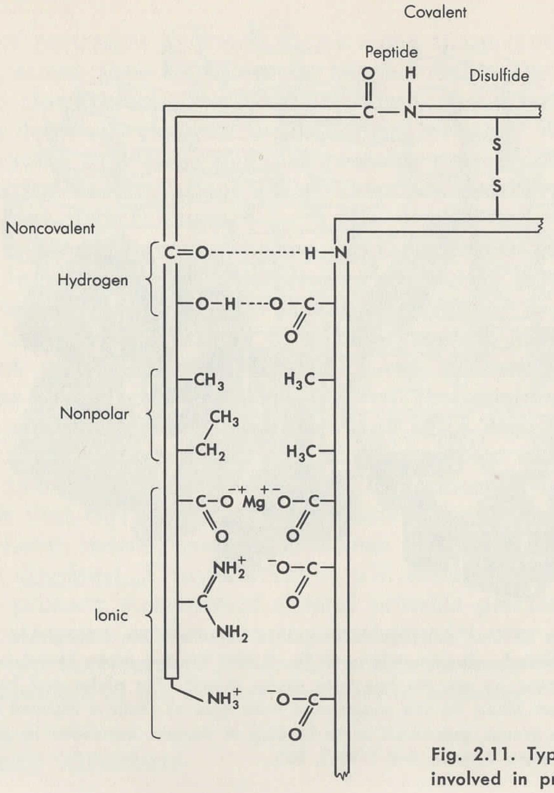


Fig. 2.11. Types of chemical bonds involved in protein structure.

the proline residue dictates that the two carbons nearest the imino peptide bond remain coplanar with the atoms comprising the bond. These two carbons of the proline ring, by protruding, sterically block formation of the alpha-helical configuration. Therefore, in the vicinity of proline residues, the polypeptide chain is unable to form an alpha helix and is forced to twist or bend. X-ray diffraction shows that all of the bends in polypeptide chains cannot be accounted for simply by the occurrence of proline residues. There must be other forces that overcome the tendency for the formation of the alpha helix and that permit the protein to assume other configurations.

One of the strong stabilizing forces in tertiary protein structure is the disulfide bond, shown at the top of Fig. 2.11. The disulfide bond (—S—S—) is a covalent linkage formed by the oxidation of the sulfhydryl groups (—SH) of two cysteine residues. Re-examination of Fig. 2.2 will demonstrate that cysteine is the only amino acid with an exposed sulfhydryl group found in proteins. In a native protein containing several cysteine residues, only certain pairs interact to form disulfide bridges at specific places in the molecule. Such bridges may be broken by chemical reduction (—S—S— → —SH + —SH) or by oxidation of the cystine residues to cysteic acid residues (—S—S— → $2 \cdot SO_3H$).

Although disulfide bridges are important in stabilizing the protein in its native configuration, other forces contribute strongly to the proper folding of the polypeptide into a three-dimensional protein molecule, as is strikingly illustrated by the following type of experiment, which has been performed with several enzymes. The disulfide bridges of the native protein are chemically reduced. The protein is then allowed to reoxidize slowly in very dilute solution. During this slow reoxidation, the disulfide bonds reform and interchange, until ultimately the native protein configuration is again attained.

One of these other forces is the hydrogen bond. Hydrogen bonding, depicted in Fig. 2.11, may take place between the C=O and H—N of nearby peptide bonds of an alpha helix (compare Fig. 2.9) or between C=O and H—N that are in different segments of the polypeptide chain, thereby contributing to the over-all folding process. Hydrogen bonds may also form between amino acid side chains—between a tyrosine hydroxyl (—OH) and a carboxyl $(\underset{\substack{|\\ \mathrm{OH}}}{\mathrm{C}}{=}\mathrm{O})$ of glutamic acid, for example. Although these amino acid residues may be far removed from each other in the linear sequence of the polypeptide chain, their affinities for the sharing of the hydrogen atom bring them into proximity in the folded protein.

Other weak bonding forces are considered to play important roles in the association of different segments within the polypeptide chain. Nonpolar amino acid side chains are attracted by hydrophobic (water-excluding) interactions, just as oil droplets tend to coalesce in water. As Fig. 2.11 shows, the hydrates that are formed may be either alike or mixed. Evidence for a role of nonpolar side chains in protein structure stems from several observations. Urea, an effective denaturant in high concentrations, also disrupts hydrates of nonpolar compounds. Addition of extra nonpolar side chains to proteins—methylation of $—NH_2$ to $—N(CH_3)_2$, for example—stabilizes them.

Finally, various types of ionic bonds may help to form and maintain the folded protein configuration. Divalent cations such as magnesium ion (see Fig. 2.11) may form electrostatic bonds with two acidic side chains. And basic and acidic groups of the constituent amino acids may also interact, as shown at the bottom of the figure.

Quaternary structure

Although some proteins are composed of a single polypeptide chain, others consist of two or more interacting chains, which may be identical or different in primary structure. This specific association of subunits into a complex macromolecule is referred to as the quaternary structure. The same forces involved in forming and maintaining tertiary structure are involved in interactions of several polypeptide chains to form a quaternary structure: disulfide bridges, hydrogen bonds, nonpolar bonds, and electrostatic bonds. One of the most striking examples of such protein-protein interactions is that occurring among

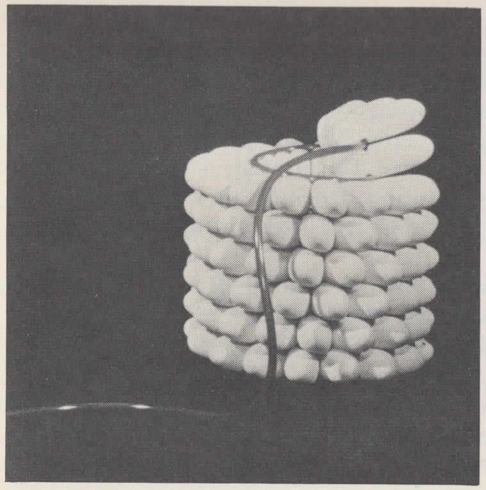
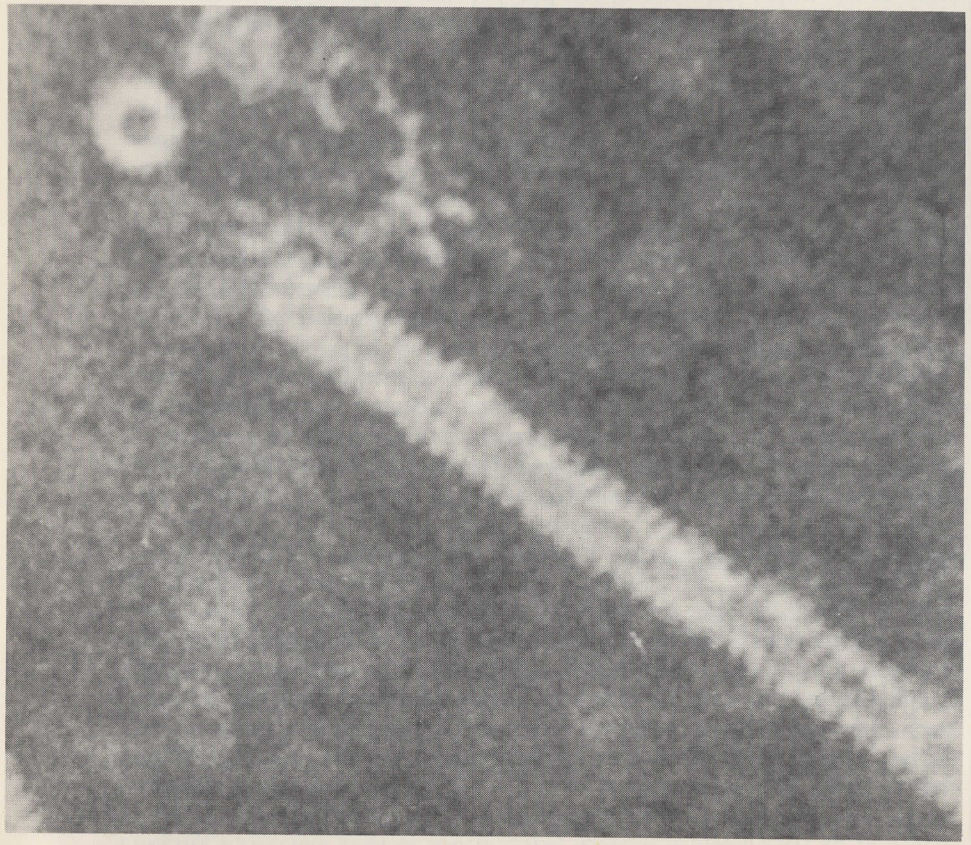


Fig. 2.12. (*a, left*) Model depicting the helical arrangement of some of the protein molecules (white) and the RNA (long gray structure) in a TMV particle. From H. Fraenkel-Conrat, *Design and Function at the Threshold of Life: The Viruses* (Academic, 1963). (*b, below*) Electron micrograph of a partially disrupted rod of TMV showing periodicity along the axis. Structures, interpreted as being individual polypeptides, are visible. From J. Nagington and R. W. Horne, *Virology, 15* (1961), 348; photograph courtesy of R. W. Horne and P. Wildy. Magnification: 720,000×.



aggregated protein subunits that form the protective outer coat for viral nucleic acids. Among such virus proteins, that of TMV has been studied most extensively.

TMV particles are rodlike, tubular structures about 3,000 angstroms long and 180 angstroms in diameter, with a molecular weight of about 40 million. The tube is constructed of a single RNA molecule 30,000 angstroms long that is coiled about the central axis of the particle. The RNA molecule is embedded in a protein matrix (the coat) that stabilizes the RNA, protecting it from disruption by changes in the enzymatic and physical conditions of the environment (Figs. 2.12a and 2.12b).

The subunits of the protein matrix have been seen in the electron microscope, and their arrangement in crystals of the virus has been studied by X-ray diffraction methods. The intact particles have been treated with carboxypeptidase, and the C-terminal threonine residues have been quantitatively recovered (see Fig. 2.4). Approximately 2,200 threonine residues are released per virus particle, indicating that the protein consists of about 2,200 polypeptide chains, all of which could be identical. The fingerprint profile of the tryptic peptides of the protein also has been used to establish the presence and identity of the polypeptide chain subunits.

Mild acid or alkali separates the RNA from the protein and fragments the protein into a number of packets of polypeptide chains. These packets readily reassociate to form long rods when the *p*H is returned to neutrality. The reconstituted rods have essentially the same structure as the protein of the initial virus rods except that, in the absence of the RNA, they form in heterogeneous lengths. They are less stable to denaturing agents than are the normal length reaggregates formed in the presence of RNA. The last three amino acids at the C-terminal end, pro.ala.thr, are not required for reaggregation and binding; they can be removed from the intact virus without noticeable effect on the stability of the rods. The polypeptide chains constituting a packet cannot be completely dissociated from one another and reaggregated into rods except under very special conditions. Many treatments that completely dissociate the polypeptide chains lead to irreversible denaturation of the protein.

These observations suggest that the orderly and specific quaternary structure of the protein depends upon critical interactions between the polypeptide chain subunits. Only when the specific tertiary structure is attained and the proper environmental conditions are provided can the appropriate interactions occur. The specific quaternary structure depends mainly on protein-protein interactions, but the limitation on rod length imposed by the presence of TMV RNA shows that accessory substances also may be involved. All evidence at present supports the view that the primary structure of the protein, which dictates the limitations of secondary and tertiary structure, also delineates the correct quaternary structure necessary for the biological activity of the molecule. Because of the specific structure of the folded polypeptide chain, only certain segments of it will be available to take part in protein-protein interactions. In TMV protein, the N-terminal tryptic peptide (Fig. 2.4) is suspected of playing a particular role in protein-protein interactions, for it tends to form aggregates of high molecular weight in vitro even when separated from the rest of the chain.

Protein configuration and activity

Cellular metabolism requires that proteins interact with other types of molecules. In some cases, these interactions can affect the tertiary structure of the protein. Indeed, an enzyme achieves its catalytic activity by binding the substrate to itself at a specific region called the active center. Enzymatic activity may be inhibited by the binding of small molecules (1) at the active center of the enzyme, so that the reaction with substrate is blocked (isosteric effects); or (2) at other positions in the protein molecule, so that its tertiary structure is changed and its affinity for substrate is subsequently altered (allosteric effects; see the discussion of feedback inhibition on pages 84-87). You will no doubt encounter this critical relationship between protein configuration and biological activity in many aspects of modern biology: in studies on the mechanism of action of hormones, the nature of the specificity of antigen and antibody reactions, and the structure and organization of multicomponent enzyme systems, to name a few.

Chapter 4 discusses the evidence that mutations affect the primary structure of proteins. From the discussion above, we already can make several predictions as to how changes in primary structure might affect certain properties of protein molecules. We might expect that substitutions of amino acids in the polypeptide could drastically affect secondary, tertiary, and even quaternary structure, thereby modifying the activity, stability, total charge, and immunological properties of the protein molecule. The extent of the change in properties should depend not only upon the site of the substitution but also upon the type of amino acid substituted. For example, replacement of a glutamic acid (with a negatively charged side chain) by a lysine (with a positively charged side chain) might be expected to have a much more drastic effect than would the substitution of a valine by an isoleucine residue. These suppositions lead us to believe that qualitative changes in the primary structure of proteins can lead to quantitative changes in the properties of proteins.

The increasingly convincing evidence that many aspects of the secondary and tertiary structure are a consequence of the primary structure has further implications for genetics. We consider that the folding and the configuration of native protein molecules, including their quaternary structure, are largely dictated by the interactions of critically positioned side chains of the amino acids composing the polypeptide chain. That this is universally true remains to be determined. Nevertheless, evidence from a few proteins is strong enough to allow us freedom, in terms of genetic considerations, to disregard the details of the folding process and to assume that folding occurs spontaneously and automatically once the amino acid sequence is determined. We can, therefore, now ask how the primary structure is determined and how mutations affect the primary structure. These subjects will be discussed in Chapters 3 and 4, respectively.

References

Anfinsen, C. B., *The Molecular Basis of Evolution.* New York: John Wiley & Sons, Inc., 1963. An inexpensive paperback edition of an excellent, simple introduction to proteins and their significance in genetics, first published in 1959.

———, "General Remarks on Protein Structure and Biosynthesis" in *Informational Macromolecules,* ed. H. J. Vogel, V. Bryson, and J. O. Lampen. New York: Academic Press, Inc., 1963. A discussion of the folding of polypeptide chains and their biosynthesis.

Bamford, C. H., A. Elliott, and W. E. Hanby, *Synthetic Polypeptides: Preparation, Structure and Properties.* New York: Academic Press, Inc., 1956.

Kendrew, J. C., "Myoglobin and the Structure of Proteins," *Science, 139* (1963), 1259. Nobel prize address.

Neurath, H., and K. Bailey, eds., *The Proteins: Chemistry, Biological Activity and Methods,* 2nd ed. New York: Academic Press, Inc., 1964.

Perutz, M. F., "X-Ray Analysis of Hemoglobin," *Science, 140* (1963), 863. Nobel prize address.

Stahman, M. A., ed., *Polyamino Acids, Polypeptides, and Proteins.* Madison: University of Wisconsin Press, 1962.

Questions

2.1. In the library, locate an original article or a review article that gives the complete primary structure of a polypeptide chain of a natural protein (e.g., ribonuclease, hemoglobin, cytochrome *c*, albumin, etc.). Note the name of the protein, its source (tissue and organism), the number of polypeptide chains it contains, and the source from which you obtained your information. Copy the amino acid sequence for the first 75 residues, starting at the N-terminal end.

(*a*) Would you find an N-terminal amino acid released from this protein by treatment with aminopeptidase?

(*b*) How many peptide fragments would you expect to obtain after digestion of the 75-residue portion of the polypeptide chain with trypsin? After chymotryptic digestion?

(*c*) Would you expect any nonpeptide products to be released by tryptic digestion? By tryptic digestion followed by chymotryptic digestion?

(*d*) How would you determine the sequence of amino acids in tryptic peptide I, the N-terminal tryptic peptide? Write out the sequence.

(*e*) Following the form of Fig. 2.1 and the structures of amino acids shown in Fig. 2.2, draw out the structure of peptide I, the N-terminal peptide, through the first four amino acid residues.

(*f*) Which amino acids in tryptic peptide I might play roles in determining the tertiary structure of the protein? Describe the role of each.

(*g*) Do any of the tryptic and chymotryptic peptides overlap in a manner that allows their unequivocal placement relative to one another solely on the basis of amino acid content (that is, lacking knowledge of the order of amino acids within the fragments)?

(*h*) Now, suppose that you know the N-terminal and C-terminal amino acids in each fragment. Again see how many fragments you can place in sequence.

2.2. Hemoglobin A is the normal, adult human hemoglobin, an oxygen-transporting blood protein. The hemoglobin molecule is composed of four polypeptide chains.

(*a*) What types of bonds do you believe might be involved in holding the four subunits of the hemoglobin protein together? What tests could you apply to differentiate some of these possibilities?

(*b*) Of the four polypeptide subunits of the hemoglobin A protein, two polypeptide chains are identical alpha chains; the other two polypeptide chains, beta chains, are different in primary structure from the alpha chains but are identical to each other. However, many of the amino acid sequences in the alpha chain are similar to those in the beta chain. What bearing do you think these facts might have on the evolution of these two kinds of polypeptide chains?

(*c*) Would you expect the alpha chains and the beta chains to fold in identical, similar, or extremely different configurations? Why?

Three

Protein Synthesis

In Chapter 2, we concluded that the structure and function of proteins are direct consequences of the number and arrangement of amino acids in linear polypeptide chains. Chapter 4 will examine evidence that the genetic material exerts its control of the phenotypic potential of cells and organisms by dictating the structure of proteins. These two concepts are enormously simplifying ones. Armed with them, we can focus attention exclusively on how the amino acid sequences are determined by the genetic material and thereby gain insight into important phases of gene action.

We are faced, then, with the question of how the correct amino acid gets built into exactly the proper position in the polypeptide chain. Although this question cannot yet be fully answered, a series of exciting developments during the last ten years allows us to sketch out some of the important steps in the process of protein biosynthesis. Knowledge of these steps allows us partially to resolve the problem of the manner in which the genetic material exercises its control of the phenotype. In other words, we are beginning to understand how genetic information stored in nucleic acid is translated and expressed in the structure and function of specific protein molecules.

Amino acid activation

The first step in protein synthesis is *amino acid*

Fig. 3.1a. ATP (adenosine triphosphate), a nucleoside triphosphate, showing structure and the numbering nomenclature for the purine ring and the ribose moeity (the latter are listed as 1′, 2′, etc., to differentiate them from the ring carbons of the base).

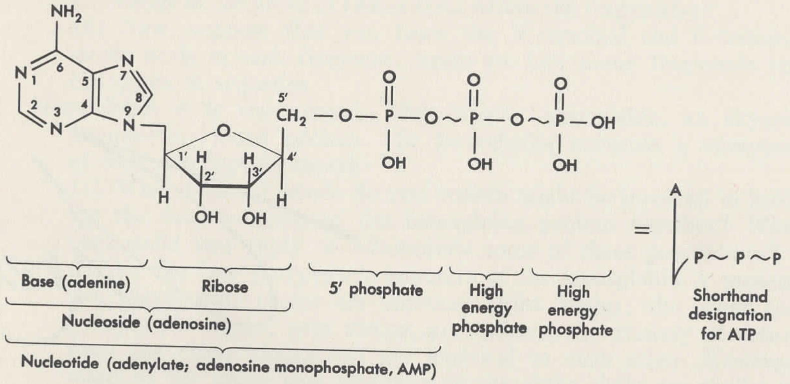
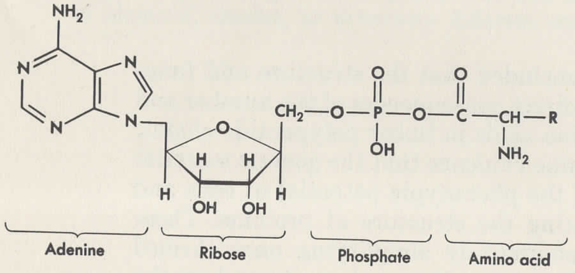

NH₂
N
CH₂—O—P—O∼P—O∼P—OH
OH OH OH
A
P∼P∼P
Base (adenine)
Ribose
5′ phosphate
High energy phosphate
High energy phosphate
Shorthand designation for ATP
Nucleoside (adenosine)
Nucleotide (adenylate; adenosine monophosphate, AMP)

Fig. 3.1b. Amino acid adenylate. R designates the side chain for any alpha amino acid.


NH₂
CH₂—O—P—O—C—CH—R
OH
NH₂
Adenine
Ribose
Phosphate
Amino acid

activation. It involves reaction of an amino acid with adenosine triphosphate (ATP); the reaction is catalyzed by a specific *amino acyl RNA synthetase.* The first product is an amino acid adenylate in which is conserved the energy required for the second step, the attachment of the amino acid to a specific sRNA molecule. The adenylate remains bound to the enzyme until it is transferred to the sRNA, a transfer that is catalyzed by the same enzyme. The structures of ATP and of an amino acid adenylate are shown in Fig. 3.1. Reaction 1 in Fig. 3.2 depicts the activation of amino acid.

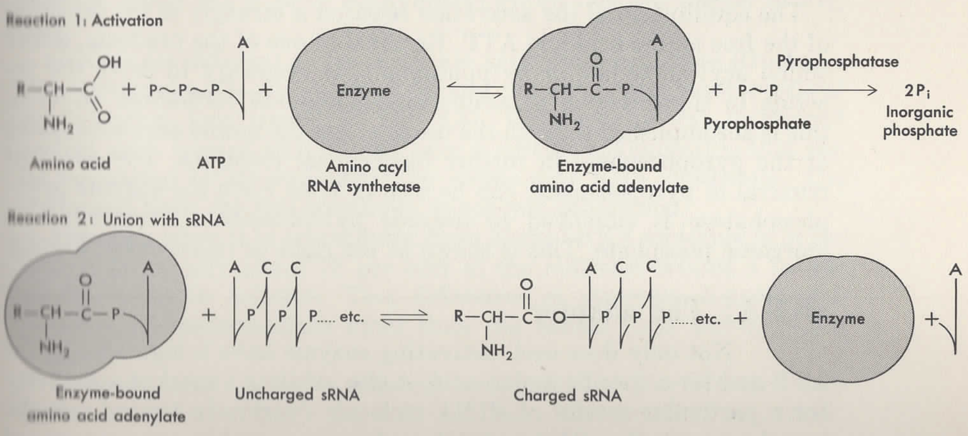
The synthetases are highly specific under physiological conditions; i.e., each of them usually activates only one kind of amino acid. Most cells normally contain relatively low levels of free metabolites of small molecular weight, together called a pool. However, cells are unable to maintain pool levels precisely enough to prevent an occasional imbalance in concentrations of the different metabolites composing the pool. Theoretically, because the activating enzyme does not have absolute specificity for its substrate, an imbalance in the amino acid pool might occasionally allow it to react with a "wrong" amino acid; such an error could cause chaos in protein synthesis. Fortunately, a crucial control is built into the specificity of an activating enzyme: its affinity for its natural, specific substrate is much higher than for any other. For example, the isoleucine-activating enzyme has a hundred fold higher affinity for isoleucine than it does for valine; even in the presence of abnormally high concentrations of valine, it will preferentially activate any isoleucine present, rather than its competitor, valine. The same situation holds for the activation of D and L amino acids; the relatively high affinity of the enzymes for the L isomers allows for their preferential activation, even in the presence of D amino acids.

There are still further checks on these enzymes. As we shall see in the next section, if the wrong amino acid is activated by an activating enzyme, the enzyme specificity is such that the amino acid adenylate in most cases is unable to take part in the next step of protein synthesis.

We have said that the adenylates remain bound to their activating enzymes (see reaction 1 in Fig. 3.2). And this is fortunate, for they are very reactive when free in solution: they form peptide bonds at random with any amino acid present, or even with amine groups on the side chains of amino acids in proteins. This high reactivity would be detrimental to specific protein synthesis, which requires a particular sequence of amino acids. The bound adenylate is thus protected from reacting nonspecifically. The configuration of the activating enzyme is such that the bound adenylate can react only with the sRNA molecules specific for its particular amino acid (see "sRNA: The adaptor").

Amino acid activation is a reversible reaction (Fig. 3.2). It is conveniently observed in the reverse direction by adding radioactive pyrophosphate to the amino acyl adenylate-enzyme complex and following the incorporation of radioactive phosphorus into ATP. The

Fig. 3.2. Amino acid activation and transfer to sRNA. R denotes an amino acid side chain. The shorthand designation for ATP, a 5′ nucleoside triphosphate, is that symbolized in Fig. 3.1a. The configuration of the amino acid-acceptor portion of the sRNA molecule follows the same shorthand convention. The nucleotides adjacent to A in the sRNA are cytidylates, designated as C.


Reaction 1: Activation
R—CH—C(OH)=O
NH₂
P∼P∼P
A
Enzyme
R—CH—C—P
NH₂
P∼P
Pyrophosphatase
2Pᵢ
Amino acid
ATP
Amino acyl RNA synthetase
Enzyme-bound amino acid adenylate
Pyrophosphate
Inorganic phosphate
Reaction 2: Union with sRNA
A C C
P P P..... etc.
R—CH—C—O
Enzyme
Enzyme-bound amino acid adenylate
Uncharged sRNA
Charged sRNA

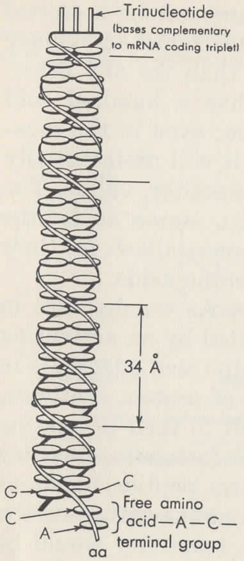

Trinucleotide
(bases complementary to mRNA coding triplet)
34 Å
G
C
A
aa
C
Free amino acid—A—C— terminal group

Fig. 3.3. Proposed general structure for an sRNA molecule. A single polynucleotide chain interacts with itself by forming a bend near the middle. Hydrogen bonding takes place between the nucleotides from the two halves of the chain and results in the formation of a double-helical structure (adenine-uracil and guanine-cytosine base pairs). The guanine from one end of the chain forms a base pair of the Watson–Crick type with the cytosine third from the terminal end. This leaves two nucleotides on the end bearing the amino acid in the flexible state. The molecule is composed of about 70 nucleotides with an aggregate molecular weight of about 24,000. Its sedimentation constant is about 4s. sRNA comprises about 5 per cent of the RNA of the cell.

sRNA molecules occur in both active and inactive forms. Inactive molecules lack the —C—C—A terminal sequence, which can be added, converting the molecules to active form, by reaction with an enzyme system that utilizes ATP and cytidine triphosphate (CTP) as substrates. Under certain conditions the —C—C—A terminal group can be detached from an active sRNA molecule enzymatically.

When the sRNA molecule is carrying an amino acid, detachment of —C—C—A is inhibited; this indicates that the amino acid is attached somewhere on this terminal sequence. More conclusive evidence for the terminal attachment stems from the observation that the enzyme ribonuclease digests sRNA, leaving among the products, amino acid–adenosine–cytidylate. From the known specificity of ribonuclease (which splits RNA at the pyrimidine 3′ phosphate linkages), it is inferred that the amino acid is attached to the terminal A.

Several unusual bases have been detected in sRNA. Most of them are methylated derivatives of those normally found in RNA. The modification of the bases appears to be carried out by specific enzymes after the formation of the polynucleotide chain. The modified bases are (in order of their frequency found in bulk sRNA): pseudouridine, 5-methylcytosine, 6-aminopurine, 6-diethylaminopurine, methylaminohydroxypurine, dimethylaminohydroxypurine, 5-methyluracil (thymine), 6-methylaminopurine. Redrawn from G. Zubay, *Science, 140* (1963), 1092.

ATP can be separated from other components in the reaction mixture, and its radioactivity can be measured.

The equilibrium of the activation reaction is strongly in the direction of the free amino acid and ATP. Removal of one of the products, either amino acyl adenylate or pyrophosphate, is necessary to drive the reaction in the direction of amino acyl–adenylate formation. In vivo, this is accomplished through the use of either the amino acid adenylates or the pyrophosphate in further biochemical reactions. The effect of removal of pyrophosphate can be seen in vitro when the enzyme pyrophosphatase is employed to degrade pyrophosphate irreversibly to inorganic phosphate. This is shown at the right of reaction 1 in Fig. 3.2.

sRNA: The adaptor

Not only does each activating enzyme have a high affinity for ATP and for a specific amino acid, it also exhibits a marked specificity for a particular species of sRNA molecule (variously termed soluble, transfer or adaptor RNA). sRNA molecules probably are synthesized at particular regions of the genetic material, each species of sRNA molecule possibly reflecting a complementarity in base composition to its particular gene in the DNA.

The dual specificity of activating enzymes allows them to carry out both the first step (amino acid activation) and the second step (transfer of amino acid to sRNA) of protein biosynthesis. As shown in Fig. 3.2, reaction 2, the second step is a reaction between enzyme-bound amino acyl adenylate and a specific sRNA. The amino acid-sRNA complex is then released from the enzyme. The sRNA is now said to be charged with amino acid and is referred to as charged sRNA.

The rate of amino acid–adenylate formation is much more rapid than amino acid–sRNA formation; therefore, the transfer reaction to sRNA is ordinarily the rate-limiting one in the over-all process. The specificity of the activating enzyme is greater in the transfer reaction than in the activation reaction. The isoleucine-activating enzyme, for example, can form a valine adenylate if isoleucine is absent and valine is present at high concentrations. However, valine adenylate, when bound on the isoleucine-activating enzyme, is not transferred either to isoleucine-specific sRNA or to valine-specific sRNA.

A model of an sRNA molecule is shown in Fig. 3.3, and some of the properties of sRNA are given in the accompanying legend. Most, if not all, active sRNA molecules possess the terminal trinucleotide cytidylic-cytidylic-adenylic. The amino acid is attached to the terminal adenylic acid residue of the sRNA chain. Except for the terminal nucleotide, the 3′ and 5′ hydroxyls of each nucleotide are covalently bound to adjacent nucleotides by phosphodiester linkages. When consulting Fig. 3.1, you will see that nucleotides have hydroxyl groups at the 2′ and 3′ positions. Compounds that possess hydroxyl groups in a *cis*-configuration are sensitive to periodate oxidation, while nucleotides that lack free *cis* hydroxyls are resistant to oxidation by this reagent. It is known that the terminal nucleotide of sRNA, when charged with an amino acid, is resistant to periodate oxidation. Consequently, the amino acid is thought to be attached to the terminal adenosine of sRNA through either the 2′ or 3′ *cis*-hydroxyl group (see the legend to Fig. 3.3).

sRNA's for different amino acids are very similar in gross structure: size, terminal —C—C—A residues, and predominantly double-helical configuration (see legend to Fig. 3.3). However, as was mentioned earlier, each type of sRNA must have a specific configuration that can be recognized by its own particular amino acyl RNA synthetase. The structural basis for this specificity is not yet understood.

Although when it is crystallized most of the sRNA molecule is in the helical form, only about 75 per cent of the molecule retains a helical configuration in solution. This difference in structure suggests the presence of conformations other than the helical form shown in the diagram.

Different sRNA's have different compositions. Furthermore, two different leucine sRNA's from *E. coli* have the terminal sequences: —G—C—A—C—C—A—leucine and —G—U—A—C—C—A—leucine.

It is possible that developments in the near future using chemical methods and X-ray diffraction techniques will allow the complete structure of pure classes of sRNA molecules to be visualized. These studies will assist in the recognition of the means by which the sRNA molecules are selected out of the pool by specific amino acyl RNA synthetases. Also, it is believed that somewhere in the sRNA molecule is a short nucleotide sequence that represents the exact Watson-Crick base complement of the mRNA code for a particular amino acid. In the model structure for sRNA shown in Fig. 3.3, this complement is shown as the short sequence at the base of the molecule (or approximately at the center of the polynucleotide chain).

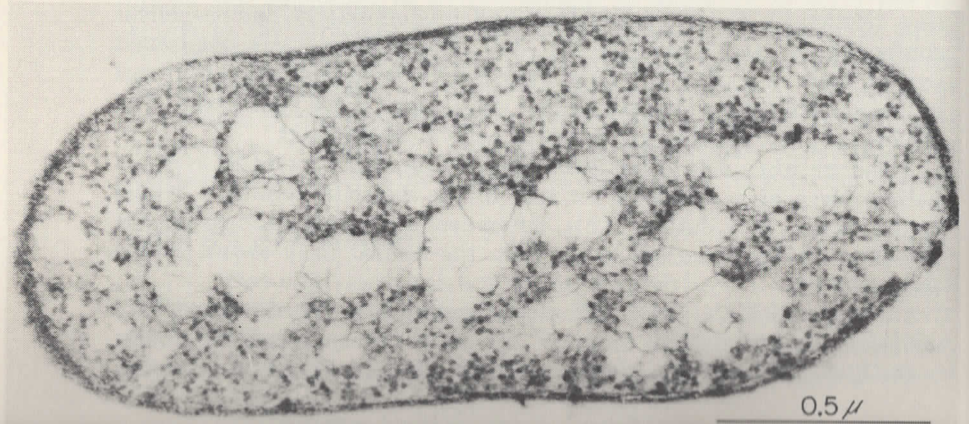
To recapitulate, the amino acid attachment sites of sRNA molecules are the same for most, if not all, classes of sRNA's. sRNA molecules contain some structural characteristics unique to each class and specific for a given species of amino acyl RNA synthetase. Finally, sRNA molecules are assumed to contain a unique short sequence of nucleotides that allows them to seek out the proper complementary series of bases on the mRNA template. Later in this chapter we will review some of the evidence that indicates the protein-forming system uses the amino acid-sRNA complexes as the raw material from which to build polypeptide chains.

Ribosomes

The correlation between the rate of protein synthesis and the RNA content of various cells reflects the presence of large numbers of ribosomes in actively growing, protein-synthesizing cells.

Ribosomes are small particles, not resolved in the light microscope, that contain about 60 per cent RNA and 40 per cent protein. In bac-

Fig. 3.4. Electron micrograph of a section through a cell of *Escherichia coli.* A bacterium is surrounded by a cell wall and, interior to this, by a cell membrane. The cytoplasm is packed with ribosomes. The DNA is contained in the lighter (less dense) portions of the cell and is not separated from the cytoplasm by a nuclear membrane. From S. F. Conti and M. E. Gettner, *J. Bacteriol., 83* (1962), 544.



teria, ribosomes are packed throughout the cytoplasm (Fig. 3.4). In certain cells of higher organisms, some ribosomes are found free and others are found attached to the surfaces of tubular lipoprotein membranes (collectively called endoplasmic reticulum). When the cell is broken open, portions of the membranes with ribosomes attached are released. The membranes can first be sedimented, together with the free ribosomes, into a pellet called the microsome fraction, and then dissolved with detergents to release the ribosomes. Ribosomes purified by centrifugation from a cell extract are shown in Fig. 3.5.

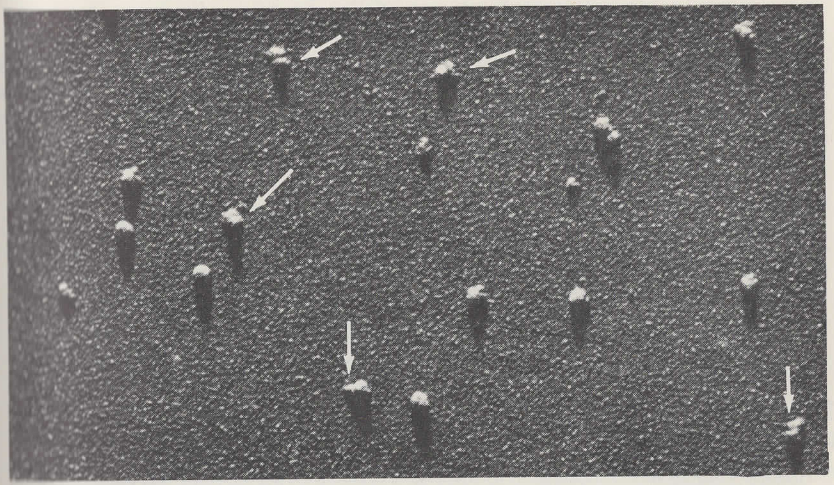


Fig. 3.5. Electron micrograph of 70s ribosomes from *Escherichia coli.* The 70s particles, indicated by arrows, show a complex structure, indicative of the union of 30s and 50s components. A few free 30s and 50s particles also are present in the preparation. Magnification: 100,000×. From C. E. Hall and H. S. Slayter, *J. Mol. Biol., 1* (1959), 329.

The ribosomes are classified by their sedimentation constants (s), determined by their rates of sedimentation in an ultracentrifuge. Ribosomes from most types of cells fall into one or more of the general classes depicted in Fig. 3.6. The ribosomes of most organisms dissociate or reassociate into specific complexes depending upon the magnesium ion concentration in the suspending medium; the concentrations shown in Fig. 3.6 apply to ribosomes extracted from cells of the bacterium *Escherichia coli.*

Ribosomal RNA differs in size and base content from sRNA (compare legend of Fig. 3.3 with Fig. 3.6). The two classes (16s and 23s) of ribosomal RNA's have base sequences that appear to be complementary to regions in the DNA along which they are presumably synthesized. In *E. coli,* the rate of ribosomal RNA formation is about one molecule per second per gene (and each RNA molecule is over 1,000 nucleotides

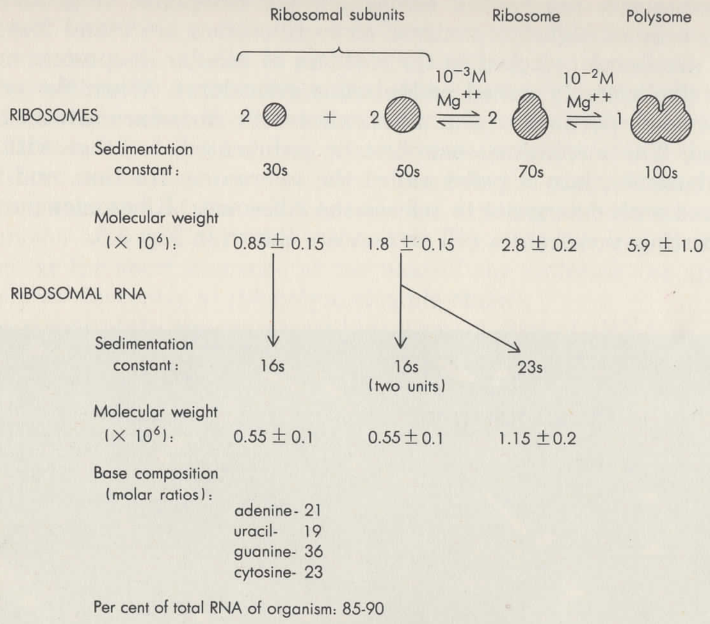

Ribosomal subunits
Ribosome
Polysome
10^{-3}M
Mg^{++}
10^{-2}M
Mg^{++}
RIBOSOMES
2
+
2
2
1
Sedimentation constant:
30s
50s
70s
100s
Molecular weight (× 10^6):
0.85 ± 0.15
1.8 ± 0.15
2.8 ± 0.2
5.9 ± 1.0
RIBOSOMAL RNA
Sedimentation constant:
16s
16s (two units)
23s
Molecular weight (× 10^6):
0.55 ± 0.1
0.55 ± 0.1
1.15 ± 0.2
Base composition (molar ratios):
adenine- 21
uracil- 19
guanine- 36
cytosine- 23
Per cent of total RNA of organism: 85-90

Fig. 3.6. Properties of ribosomes.

long!). The ribosomal RNA then combines with proteins (in some unknown manner) to form a 30s or 50s component. Once protected by combination with protein, ribosomal RNA is stable and is not readily degraded in vivo. In mammalian cells, synthesis of ribosomes appears to be a more complex process in which high molecular weight (45s) RNA is a precursor of ribosomal RNA.

The reason for the complex structure and the RNA content of ribosomes is not yet clear. The structural RNA of ribosomes lacks the specificity required for the precise designation of the sequence of amino acids needed to build a specific polypeptide chain. Rather, the ribosomes appear to be relatively nonspecific workbenches: that is, any one of a number of proteins can be made on a given ribosome.

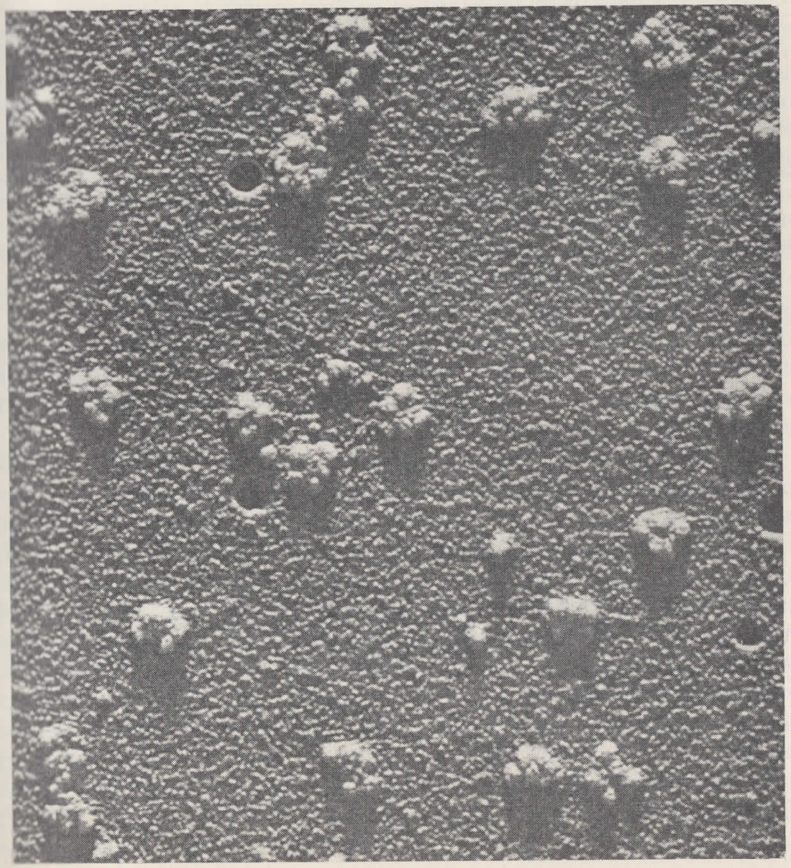
The ribosomes active in protein synthesis are the 70s ribosomes. Aggregates of 70s ribosomes are called polysomes. Only 100s polysomes are shown in Fig. 3.6, but much larger aggregates are known (some are shown in Fig. 3.7). The 70s particles of the polysome seem to be held together by mRNA molecules adsorbed to the 30s components (Fig. 3.8).

There is evidence for several high-affinity attachment sites for sRNA on the 50s component of the ribosome. In ribosomes active in protein synthesis, an sRNA molecule, attached by ionic bonds to one of the sites, is itself covalently linked with an incomplete polypeptide chain

in the process of synthesis (Fig. 3.8). The attachment of sRNA to ribosomes does not require an enzyme, nor does it require that the sRNA be charged with an amino acid. It does require the presence of the —C—C—A end and magnesium ion. There are some 20 to 50 additional sites to which sRNA can bind with lower affinity; the binding is facilitated by the presence of mRNA complexed to the ribosome.

We have, then, the ribosomal workbenches to and from which sRNA molecules can shuttle, carrying amino acids for the synthesis of the polypeptide chain. The specific order in which these charged sRNA molecules are utilized, however, depends upon the code contained in mRNA molecules, also bound to the ribosomes.

Fig. 3.7. Polysome clusters of 70s ribosomes. These organelles are active in protein biosynthesis and are believed to be held together by long mRNA molecules. Those shown in the electron micrograph were obtained from reticulocytes synthesizing the protein hemoglobin. Most of the reticulocyte polysomes are made up of five 70s ribosomes. Magnification: 100,000×. From J. R. Warner, A. Rich, and C. E. Hall, *Science, 138* (1962), 1399.



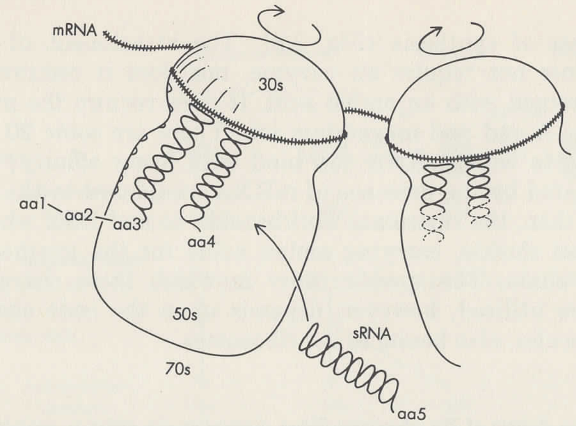


Fig. 3.8. One portion of polysome active in protein biosynthesis. Each 70s ribosome component is pictured as rotating as the code in the mRNA is successively read.

mRNA: The messenger

Proteins are built upon mRNA templates. Current evidence indictates that mRNA molecules are ribopolynucleotide complementary copies of one strand of the double-stranded DNA molecules. After its synthesis in the nucleus, mRNA becomes a component of the cytoplasm and is found associated with the ribosomes.

mRNA is synthesized rapidly; it is estimated that one molecule is synthesized per second per gene in bacteria! But the mRNA of bacteria and some mammalian cells is unstable: its half life is only a few minutes. This is in contrast to the relative stability of sRNA and ribosomal RNA. About four-fifths of the RNA synthesized in bacteria is mRNA, but, because it is degraded so rapidly, it constitutes only a few per cent of the total RNA. mRNA of other cells, for example reticulocytes, is much more stable. Thus, the stability of mRNA appears to depend upon the cell type in which it is present.

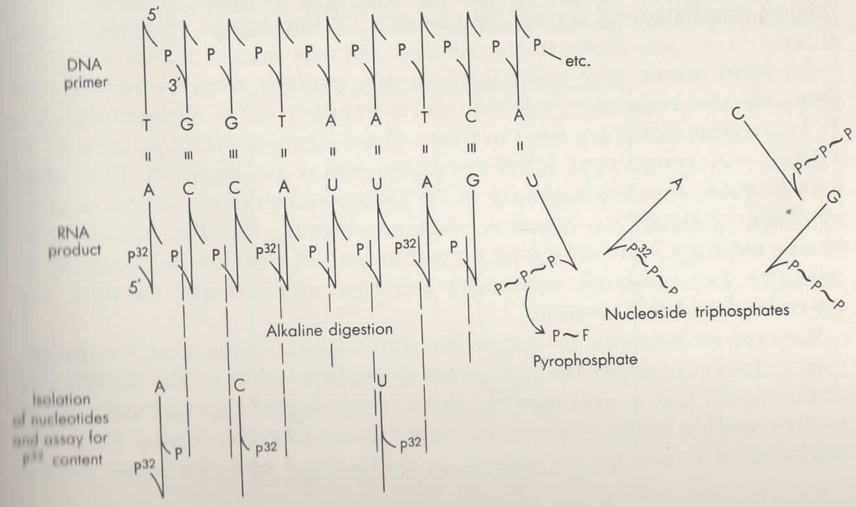
RNA with the properties of mRNA has been synthesized in vitro by an enzyme called RNA-polymerase. RNA-polymerase action requires the presence of a DNA polymer as primer. This suggests that the enzyme preferentially synthesizes RNA complementary to the DNA rather than joining nucleotides in a random fashion. Supporting this idea is the requirement of the presence of all four ribonucleoside triphosphates—ATP, guanosine triphosphate (GTP), cytidine triphosphate (CTP), and uridine triphosphate (UTP)—for maximal RNA synthesis in the presence of DNA primer. All four ribonucleotides are incorporated in the RNA product. If the primer is deoxy-A—T copolymer, however, only UTP and ATP are used for synthesis of the U-A ribonucleotide polymer, and if it is deoxy-T homopolymer, only ATP is used for synthesis of the A ribonucleotide polymer.

Although some nucleic acids isolated from different organisms are alike in gross base content, they usually differ in the average frequencies with which particular bases are found in juxtaposition along the chain. These frequencies are called nearest neighbor frequencies, and the method for their determination in RNA is shown in Fig. 3.9. Data from experiments using this technique have shown that the RNA product has a structure complementary in base sequence to the DNA primer. Furthermore, in vitro some of the product of RNA polymerase action is a hybrid of complementary RNA and DNA chains. The chains are wound about each other and thus are relatively resistant to digestion either by deoxyribonuclease (which digests free DNA) or by ribonuclease (which digests free RNA).

When radioactive phosphorus is presented to cells, sampling at short intervals of time shows that much of it enters their mRNA. The radioactive label is found first in the nucleus and later in the cytoplasm, bound to the ribosomes. Since the mRNA is not tightly bound to ribosomes, as is the bulk of the RNA (ribosomal RNA), it

Fig. 3.9. Determination of nearest neighbors in RNA synthesized along a DNA primer. A portion of the base sequence of the effective DNA (primer) strand is shown on the top row. All four nucleoside triphosphates (ATP, UTP, GTP, and CTP) are required for RNA synthesis. Only the ATP in the phosphate adjacent to the sugar is labeled with radioactive phosphorus (P^{32}). Recall that the free nucleoside triphosphates are 5′ phosphate esters (see Fig. 3.1a).

In the presence of RNA-polymerase, RNA is synthesized by addition of the 5′ nucleotides and release of the two terminal phosphates as pyrophosphate (center row in figure). As you can see, the standard base-pairing rules apply. The RNA product is isolated and subjected to alkaline hydrolysis. Alkali splits all of the 5′ phosphate ester linkages in the RNA. The nucleotides are separated, and the presence of radioactivity is determined. The 5′ radioactive phosphate, originally esterified to the adenine-ribose moiety, is now isolated as the 3′ ester of the nucleoside to the left of adenosine. The amount of radioactive phosphorus in each nucleotide gives an accurate estimate of the frequency with which that base occurs next to adenine in the newly synthesized RNA. The experiment can be repeated with the radioactive phosphorus in each of the other nucleoside triphosphates.



can be separated from the ribosomes and sRNA by dissociation and differential centrifugation. The mRNA fraction is composed of nucleotides that mimic the DNA in base composition, or, in cases where the base compositions of the two strands of DNA can be differentiated, it is found that the mRNA made in vivo has a base composition precisely complementary to only one of the two strands of DNA.

mRNA has an affinity for hybrid formation only with DNA from the same or from genetically very closely related organisms. The formation, in vitro, of such specific RNA-DNA hybrids first requires that the DNA be denatured by heating, so that the hydrogen bonds between the Watson–Crick base pairs are disrupted. If the DNA is then quickly cooled, few of the hydrogen bonds reform, leaving a collapsed, tangled molecule whose bases are free to pair, by hydrogen bonding, with added mRNA. This important experimental test for complementarity is discussed further in Chapter 7.

Not all mRNA is copied along DNA templates. Certain viruses—for example, tobacco mosaic virus—contain RNA instead of DNA as the genetic material. This RNA can serve directly as mRNA. It also can be replicated in the host cell during virus growth through the mediation of an enzyme that synthesizes a complementary RNA chain in the presence of the *RNA primer.* This complementary chain is then presumed to serve as template for its opposite member, the RNA chain found in mature virus particles and active as messenger.

The protein-forming system

Assay for amino acid incorporation

The incorporation of radioactive amino acids into protein is conveniently followed by precipitating the proteins and larger polypeptides with acid. The amino acids and small peptides remain soluble. The amount of radioactive amino acid that becomes *acid insoluble* during incubation is an initial measure of the extent of protein synthesis.

In vitro amino acid incorporation into protein requires various cell constituents: ribosomes, mRNA, and a complete set of charged sRNA's. If free amino acids are used in place of the charged sRNA's, then ATP, sRNA, and amino acyl RNA synthetases also are required. The addition of guanosine triphosphate (GTP) stimulates incorporation in some systems, although its function remains obscure. Finally, protein synthesis requires another enzyme system, called peptide polymerase, to catalyze peptide-bond formation between amino acids as they are assembled on the ribosome.

Several criteria can be applied to assure that amino acid incorporation indeed represents the formation of protein—that is, the building of amino acids into a polypeptide chain. Amino acid incorporation must be irreversible in the presence of excess nonradioactive amino acid. The acid-insoluble precipitate should be hydrolyzed and the tracer amino acid should be recovered and identified. Finally, amino acid incorporation should be prevented by compounds that can specifically inhibit protein synthesis, such as chloramphenicol and puromycin.

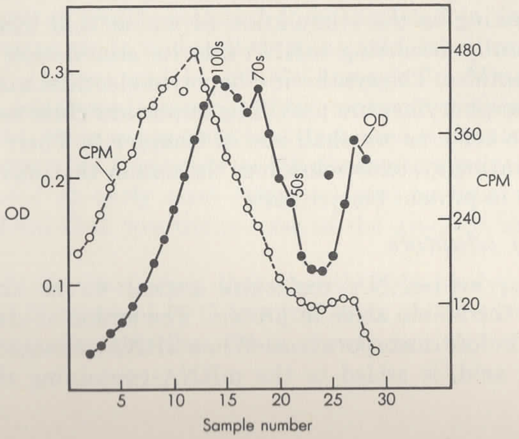
Active ribosomes

Studies of the attachment of mRNA to ribosomes and of the incorporation of amino acids into proteins in ribosomal systems often utilize the techniques illustrated in Fig. 3.10. Centrifugation through a sucrose density gradient separates materials found in cell extracts. The large particulate elements of the cell sediment most rapidly; the smaller the molecules, the nearer they remain to the top of the centrifuge tube.

In the experiment of Fig. 3.10, living cells were labeled with radioactive uracil. About a minute later, the cells were disrupted and centrifuged. Centrifugation was done in the cold in order to reduce loss of the active components by enzymatic degradation, and magnesium ion was present to lessen dissociation of ribosomes. The different species of ribosomes were located by examining each centrifuged fraction for its

Fig. 3.10. Sedimentation of radioactivity associated with pulse-labeled RNA. Bacteria infected with bacteriophages were allowed to incorporate radioactive (C^{14}) uracil for a very brief (pulse) period. They were then broken open, and the unbroken cells and debris of high molecular weight were sedimented from the extract. The clarified extract was then layered on top of a sucrose gradient in a centrifuge tube. The tube was centrifuged for several hours at high speed, the bottom of the tube was punctured with a hypodermic needle, and ten-drop samples were collected for analysis. The optical density (OD) of each sample was measured at 2,600 Å as an indication of its nucleic acid content. Each fraction was precipitated with trichloracetic acid, and its content of radioactive uracil was measured in counts per minute (CPM). The acid-precipitable radioactivity (contained in RNA) was found to be associated predominantly with a rapidly sedimenting fraction greater than 100s in size.

If a similar extract is prepared from nonradioactive bacteria, allowed to incorporate radioactive amino acids in vitro, and then centrifuged, curves are obtained for OD and CPM that are qualitatively similar to those shown here. This indicates that the fractions containing pulse RNA are the ones competent in protein biosynthesis. Redrawn from R. W. Risebrough, A. Tissieres, and J. D. Watson, *Proc. Nat. Acad. Sci. U.S., 48* (1962), 430.



absorption of ultraviolet light at a wavelength where nucleic acids strongly absorb. The small amount of newly synthesized material was traced by assaying the various samples for their radioactivity.

The results show that radioactivity rapidly enters material that sediments faster than the 100s ribosomes; this is called the heavy peak. Similar experiments show that radioactive amino acids also rapidly enter and become fixed in this same fraction. Upon treatment with ribonuclease, the radioactivity is displaced into the 70s peak. Or, exposure for several minutes to nonradioactive precursors displaces much of the label in the amino acids from the heavy peak to the 70s peak. If the various cell fractions are exposed to radioactive amino acids in vitro, only the heavy peak is active in amino acid incorporation. These experiments demonstrate that (1) it is newly synthesized RNA that is attached to ribosomes and makes them cluster into polysomes, and (2) protein synthesis is associated mainly, if not exclusively, with such polysomes. The 70s ribosomes that have completed synthesis leave the polysome cluster and are unable to participate in protein synthesis again until they become reattached to mRNA.

mRNA: Specific template

Several observations implicate mRNA and rule out sRNA and ribosomal RNA as the template in protein biosynthesis. mRNA elicits amino acid incorporation in vitro; sRNA and ribosomal RNA do not. sRNA is believed to be ineffectual because of its secondary structure —double-stranded RNA's fail to stimulate amino acid incorporation. Ribosomal RNA has some as yet unknown properties that disallow its binding to intact ribosomes, the sites of amino acid incorporation.

Furthermore, *E. coli* sRNA and activating enzymes can transfer amino acids to rabbit reticulocyte ribosomes that are synthesizing the rabbit protein, hemoglobin, in vitro. Also when f2 phage RNA, is placed with *E. coli* sRNA, ribosomes, and the necessary enzymes from *E. coli,* phage protein is synthesized. This is very strong evidence that the mRNA programs the ribosomes for the type of protein that will be produced.

Just as striking as the stimulation of amino acid incorporation by various naturally occurring mRNA's is its stimulation by synthetic ribopolynucleotides. The synthetic ribopolynucleotides made with polynucleotide phosphorylase are providing important clues as to the nature of the genetic code, as we shall see in Chapter 9. Their use in amino acid–incorporating systems reinforces the notion that mRNA plays the directive role in protein biosynthesis.

Busy adaptors

Amino acid–sRNA molecules appear to be the immediate precursors of the amino acids in protein. The amino acids do not again become free before incorporation. When sRNA, charged with radioactive amino acid, is added to the mRNA-containing ribosomes, the radioactive amino acid rapidly becomes incorporated into proteins that are first found associated with ribosomes and later released free in solution. If a great excess of nonradioactive free amino acid is added to such a system, the incorporation of the sRNA-bound radioactive amino acid into protein is not decreased. Finally, the polypeptide chains that are bound to the ribosomes contain single sRNA molecules attached to the C-terminal end of each polypeptide chain.

The sRNA molecule is not permanently altered by its reaction with the specific amino acid and the incorporation of the amino acid into the polypeptide chain. It is thus free to pick up another charge of specific amino acid for transport to the sites of protein synthesis. The fact that amino acids attached to sRNA are incorporated directly into protein without the addition of activating enzyme shows that incorporation is not mediated by the activating enzymes themselves. These events are schematically illustrated in the summary diagram of this chapter, Fig. 3.12.

Growth of the polypeptide chain

That an sRNA molecule is attached to the C-terminal end of a polypeptide chain extracted from ribosomes suggests that the chain grows linearly, starting at the N-terminal end and progressing toward the C-terminal end. Better evidence for this linear growth is derived from the following experiment (Fig. 3.11).

Rabbit reticulocytes (cells that synthesize predominantly hemoglobin) contain ribosomes with incomplete hemoglobin polypeptide chains of various lengths (time t_1 in Fig. 3.11). After various lengths of exposure to radioactive amino acids (time t_2, time t_3, time t_4), the soluble hemoglobin is separated from the ribosomes. The incomplete polypeptide chains are dissociated from the ribosomes and subjected to tryptic digestion, and the soluble hemoglobin also is treated with trypsin (the vertical lines in the figure indicate places where trypsin fragments the molecules). The various peptide fragments, whose sequence in the hemoglobin molecules is known, are then characterized and analyzed for the presence or absence of radioactivity.

The experimental results matched those depicted in the figure. At short times of exposure to radioactive amino acid, label was found near the C-terminal (finish) end in soluble hemoglobin. With increasing time, the label penetrated deeper into the molecule, with a gradient of label through the array of peptides that decreased with increasing time. The incomplete peptide fragments on the ribosomes were approximately uniformly labeled at early time (time t_1) and as time went on developed a gradient that was the reverse of the gradient in the soluble hemoglobin.

In addition to demonstrating that growth of the polypeptide chain progresses from N-terminal to C-terminal ends, this experiment is one of the best demonstrations that chain growth actually occurs on the ribosomes.

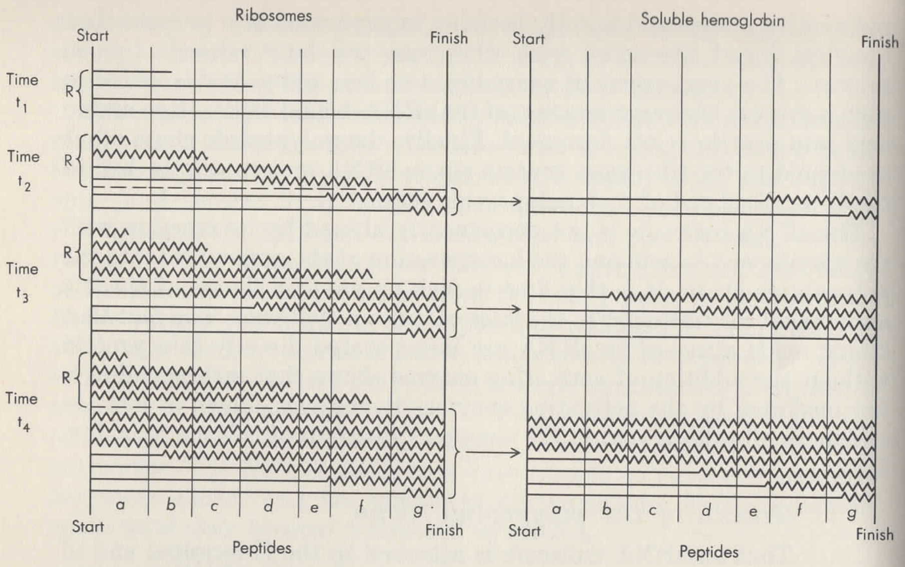

Ribosomes
Soluble hemoglobin
Start
Finish
Start
Finish
Time
t_1
Time
t_2
Time
t_3
Time
t_4
R
R
R
R
a
b
c
d
e
f
g
Start
Finish
Peptides
a
b
c
d
e
f
g
Start
Finish
Peptides

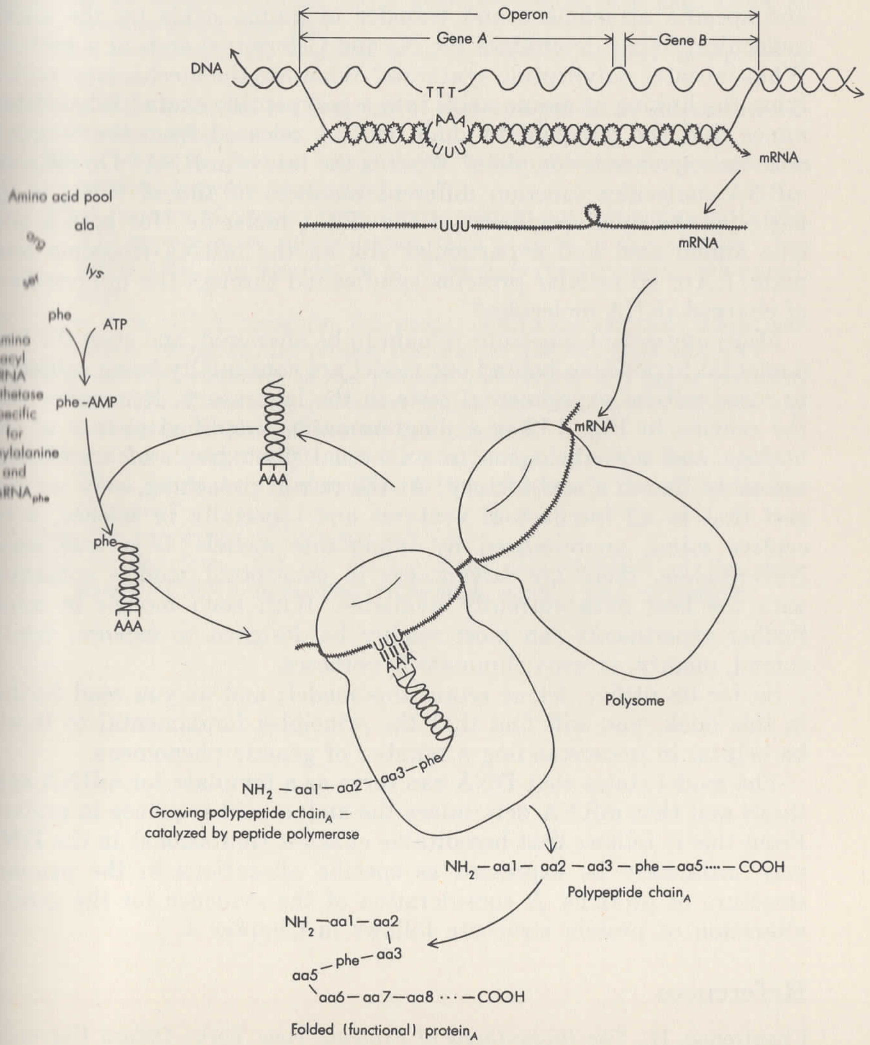
Fig. 3.11. Model of sequential chain growth from the N-terminal to the C-terminal end. The experiment is described in the text. The straight lines represent unlabeled polypeptide chains. The zigzag lines represent radioactive polypeptide chains formed after the addition of a radioactive amino acid at time t_1. The groups of peptides designated R are unfinished and remain attached to the ribosomes. The rest of the chains, having been completed, are assumed to be present in the soluble hemoglobin. In the ribosomes at time t_2, the top two lines represent peptide chains formed completely from amino acids during the time interval between t_1 and t_2. The middle two lines represent chains that have grown during the time interval but have not reached the finish line and are, therefore, still attached to the ribosomes. The bottom two chains represent those that have crossed the finish line, left the ribosomes, and are to be found mixed with other molecules of soluble hemoglobin. Redrawn from H. M. Dintzis and P. M. Knopf, in *Informational Macromolecules*, H. J. Vogel, *et al.*, eds. (Academic, 1963), p. 376.

The over-all scheme

The rate of formation of proteins is amazingly rapid. To account for the production of new hemoglobin, it has been estimated that reticulocytes must, on the average, make one polypeptide chain of hemoglobin per ribosome every minute and a half. Since there are about 150 amino acids per chain, these estimates indicate that a chain grows at a rate of about two amino acids per second!

Fig. 3.12 reviews some critical steps in protein biosynthesis as we picture them at the current time. Amino acids are selected from the intracellular amino acid pool by specific amino acyl RNA synthetases. The activated amino acids remain attached to the enzymes until they are transferred to specific sRNA molecules. The amino acid–sRNA molecules become bound to ribosomes actively engaged in protein synthesis (polysomes) and, in some way, presumably owing to the nature of Watson–Crick base pairing, are specifically oriented in the proper positions on the template mRNA. The mRNA, synthesized along one strand of the duplex DNA molecule, is bound to the ribosome in a manner that makes it accessible to the amino acid–sRNA molecules. The polypeptide chain is initiated at the N-terminal end. Amino acids are added to the growing C-terminal end of the polypeptide chain one

Fig. 3.12. A model for the genetic control of protein biosynthesis. The organization of genes into operons and the synthesis of more than one protein from a single mRNA molecule are topics discussed in Chapters 7 and 8.


Operon
Gene A
Gene B
DNA
TTT
AAA
UUU
mRNA
Amino acid pool
ala
lys
UUU
mRNA
phe
ATP
Amino acyl RNA synthetase specific for phenylalanine and $sRNA_{phe}$
phe-AMP
AAA
phe
AAA
mRNA
UUU
AAA
Polysome
NH_2 — aa1 — aa2 — aa3 — phe
Growing polypeptide $chain_A$ — catalyzed by peptide polymerase
NH_2 — aa1 — aa2 — aa3 — phe — aa5 ···· — COOH
Polypeptide $chain_A$
NH_2 — aa1 — aa2
aa5 — phe — aa3
aa6 — aa7 — aa8 ··· — COOH
Folded (functional) $protein_A$

at a time in a reaction catalyzed by an enzyme system, peptide polymerase. The specific organization and alignment of amino acids in the chain is dictated by the base sequence of the mRNA, which selects the correct complementary base sequence from the pool of charged sRNA molecules. When it is completed, the polypeptide chain folds into the final biologically active configuration and eventually is released from the ribosome.

What next?

What are the unique structural features of active ribosomes that allow adsorption and release of mRNA and that allow such rapid and specific attachment and transfer of amino acids by the sRNA molecules? What determines the N- and C-terminal ends of a protein? What initiates polypeptide synthesis? What are the mechanisms underlying the linking of amino acids into a polypeptide chain? What determines that the polypeptide chain will be released from the template once its sequence is complete? What is the fate of mRNA? Do different mRNA molecules function different numbers of times? What is the basis for the dual specificity of the sRNA molecule (for both a specific amino acid and a particular site on the mRNA-ribosome template) ? Are all cellular proteins synthesized through the intermediary of charged sRNA molecules?

Many important questions remain to be answered, and even the fundamental hypotheses behind our model are continually being subjected to more critical experimental tests in the laboratory. Remember that the scheme in Fig. 3.12 is a diagrammatic, simplified picture of our notions and not analogous to an actual photograph of an existing assembly line in a real factory! At the risk of preaching, may we suggest that in all intellectual ventures and especially in science, a receptive mind, unprejudiced by immutable models, is a true asset. Nevertheless, there are advantages to conceptual models consistent with the best data currently available. With such models in mind, further experiments can most readily be designed to explore, verify, extend, modify, or even eliminate hypotheses.

So for its utility, let us retain this model; and as you read further in this book, you will find that the principles fundamental to it will be helpful in understanding a number of genetic phenomena.

The model states that DNA can serve as a template for mRNA synthesis and that mRNA determines the amino acid sequence in protein. From this it follows that hereditable changes (mutations) in the DNA may ultimately be expressed as specific alterations in the primary structure of proteins. A consideration of the evidence for the genetic alteration of protein structure follows in Chapter 4.

References

Chantrenne, H., *The Biosynthesis of Proteins*. New York: Oxford University Press, 1961. A detailed summary of the subject with a useful guide to the original literature.

Schweet, R., and J. Bishop, "Protein Synthesis in Relation to Gene Action," in *Molecular Genetics*, Part I, J. Herbert Taylor, ed., Chap. VIII. New York: Academic Press, Inc., 1963. A good, relatively concise, and up-to-date review of the steps in protein biosynthesis with references to the original literature.

Zamecnik, P. C., "Historical and Current Aspects of the Problem of Protein Synthesis," *Harvey Lectures, 54* (1960), 256.

———, "Unsettled Questions in the Field of Protein Synthesis," *Biochem. J., 85* (1962), 257.

Questions

3.1. Draw an adenine-thymine and a guanine-cytosine base pair. Draw an adenine-uracil base pair.

3.2. What components are required in an in vitro amino acid–incorporating system?

3.3. In tabular form, list the function of each component in an amino acid–incorporating system.

3.4. What steps in protein biosynthesis utilize Watson–Crick hydrogen bonding between base pairs?

3.5. What features of protein biosynthesis pointed out schematically in Fig. 3.12 require more substantiation than is given in this chapter? Describe what you consider to be three important voids in our knowledge.

3.6. How does sRNA recognize the proper activating enzyme? How does sRNA recognize the proper position on the template mRNA?

3.7. Write a brief summary of the purpose and results of a paper reporting on original research published within the last twelve months and pertaining to the subject of this chapter.

3.8. What forces hold the 30s and 50s components of ribosomes together? What forces hold the mRNA to the ribosome? The sRNA to the ribosome?

3.9. List the properties of bacterial mRNA.

3.10. In tabular form, list the distinguishing characteristics of the three classes of RNA molecules: ribosomal, soluble, and messenger.

Four

Mutant Proteins

How do we know that the genetic material, nucleic acid, dictates the structure of protein? Increasingly numerous examples, from the genetics of man to the genetics of viruses, afford experimental confirmation of this relationship; the only known exceptions are those genes that control production of RNA's with special functions in the cell, that is, sRNA and ribosomal RNA.

Operational definitions of the gene

Stahl in *The Mechanics of Inheritance* describes the structure and properties of the nucleic acids, particularly DNA. Stahl presents the main evidence that DNA is the genetic material and that it is accurately replicated and passed to the progeny. He also describes how hereditary mutations occur in the DNA through alterations in the base composition or sequence in its polynucleotide chains. He discusses the use of inheritance tests for distinguishing mutant genetic sites located at different positions on a polynucleotide chain. The mechanism of redistributing genetic material is called *recombination.* If two mutations are located on separate chromosomes, they segregate randomly; if they are located on a single chromosome, or on a single nucleic acid molecule, they may separate with less than random frequency and are said to be linked. Generally, the closer the location of two mutations along a polynucleotide chain, the less frequent their separation, i.e., the lower the recombination frequency between the two mutations. Two independently derived mutations that affect the same base pair in the DNA molecule never show recombination.

Finally, Stahl surveys some of the evidence that the polynucleotide chain consists of segments, each controlling the expression of a particular trait. That is, one sequence of base pairs along the polynucleotide chain affects one function, whereas another sequence along the chain affects a second function.

Mutation, recombination, function: these are the three operational methods by which genetic tests delineate the structure of the hereditary material.

All three methods arose as purely genetical procedures and originally were considered to give identical estimates of the mysterious units of heredity, the genes. Later it was discovered, by examining large numbers of progeny, that some mutations in the structure presumed to represent a single gene were separated by recombination (for example, one recombinant per 10,000 progeny). The advances in this area of fine-structure genetics were tremendously facilitated through work, in the 1950's, with microorganisms whose rare recombinants could be detected and quantitatively scored much more easily than could those of higher organisms. As a result, we now recognize the gene as a section of a polydeoxyribonucleotide chain, divisible both by mutation and recombination and concerned with a specific unitary function.

One gene, one polypeptide chain

Mendel considered that a particular trait is derived from a pair of factors, one inherited from each parent, and that when the two factors differ only one is expressed: i.e., one is dominant over the other. After the rediscovery and confirmation of Mendel's basic observations on the mode of inheritance of unit factors, some geneticists hypothesized that the dominant factor is the presence of a gene; the recessive factor, its absence. However, this theory rapidly met with difficulties in interpretation when incomplete dominance (intermediate traits) and three or more slightly different alleles for a single gene were discovered. While genes were found to be located at fixed positions on the chromosomes and came to be considered as the material units of inheritance, their apparent complexity was puzzling. Because cellular physiology and the chemical nature of the genetic material were little understood at the time, it is not surprising that interpretations of gene structure and function were limited in scope and understanding.

What is known about how genes act? What does the functional genetic unit mean in biochemical terms? Answers to these questions initially were sought in the relationship between genes and specific biochemical reactions. In the early 1900's, a considerable amount of work was done on the chemistry of the anthocyanin pigments and the genetics of pigmentation in the snapdragon and other flowering plants. Several genes were found to control the onset of various identifiable and specific chemical reactions. These reactions were found to involve modifications in the structure of heterocyclic organic molecules called anthocyanins, and the reactions were found to be catalyzed by enzymes. Also in the early 1900's, Sir Archibald Garrod, a British physician interested in a number of congenital metabolic diseases in man, recognized that certain human biochemical deficiencies were caused by enzymatic abnormalities. But these very important observations, suggesting a relationship between genes and enzymes, had relatively little influence on contemporary genetic or biochemical thought.

In the 1930's, George Beadle, Boris Ephrussi, and Edward Tatum began a series of developmental genetic studies on the fruit fly, *Drosophila melanogaster*. Using a series of eye-color mutants to study the genetics of eye-pigment formation, they were able to show that the synthesis of normal eye pigment proceeded through a series of chemically identifiable intermediates and that each biosynthetic step was under distinct genetic control. From these studies, some of the first principles of biochemical genetics were formulated, some of the technical limitations of the *Drosophila* system were evaluated, and a clear-cut method of attack for studying the problem of gene-enzyme relationships in the fungus *Neurospora crassa* was established. However, until Beadle and Tatum formulated the "one gene, one enzyme" concept in the early 1940's, discussions of primary gene function remained somewhat nebulous and fluid. The theory of Beadle and Tatum states that the gene exerts its influence on the phenotype through its role in the production of an enzyme.

Research in the last two decades has verified and extended this basic conclusion and has spelled out in more detail the chemical nature of genetic material. Our picture of the structure of proteins (Chapter 2) and of the mechanism of protein biosynthesis (Chapter 3) has been formulated only in recent years. We now know that mRNA is an intermediate between the gene and the protein, and we know some of the complexities in the structure and function of enzyme molecules. Therefore, currently, the Beadle-Tatum concept of one gene, one enzyme has been revised to one gene, one polypeptide chain. The remainder of this chapter defines more clearly this relationship.

The particular examples of gene-protein relationships we have selected for discussion occur in microorganisms and were chosen because the genetic analysis has progressed to an advanced state and/or detailed analysis of the specific protein is available. The reader who thoroughly understands these examples will be able to extrapolate readily to other genetic systems. Certainly, one striking outcome of the genetics of microorganisms is the satisfactory application of concepts derived therefrom to situations in higher organisms where the genetic data may be relatively limited.

Gene-enzyme relationships

Some of the most valuable mutants isolated to date have been those that cannot synthesize one or another metabolite needed for growth and thus require particular growth factors not required by the wild-type organisms. The reasons for their utility are apparent. Most wild-type enteric bacteria grow in a synthetic medium containing inorganic salts and an organic carbon source. Wild-type *Neurospora crassa* grows on a simple medium supplemented with one of the B vitamins, biotin, and an organic carbon source. From these components in the minimal medium, the organisms can synthesize all of their protoplasmic constituents, including amino acids, purines, pyrimidines, fats, vitamins, etc.

Neurospora mutants defective in the synthesis of metabolites critical for growth are obtained by treating the asexual spores, called conidia, with a mutagen and allowing them to germinate in minimal medium. Wild-type conidia form long, threadlike structures called hyphae. The hyphae are retained by a cheesecloth filter, while the ungerminated and slowly germinating mutant conidia that cannot grow on minimal medium pass through with the medium. They are allowed to germinate on more complex medium, and their specific growth requirements are subsequently determined.

Bacterial mutants that require supplements can be concentrated in a population by using the antibiotic penicillin. Penicillin inhibits the synthesis of bacterial cell walls. When it is added to minimal medium, the wild-type bacteria grow rapidly but, without cell walls, burst open and are killed. The mutants, unable to grow on minimal medium, survive. They are transferred to a medium containing added nutrients, and then the specific growth requirement of each mutant is individually tested.

Mutations that affect a particular phenotype (for example, growth requirement for tryptophan) are localized at specific chromosomal sites through genetic tests. Localization of the genetic markers is accomplished by methods analogous to those described in Stahl's *The Mechanics of Inheritance* for mapping of the chromosomes of *Chlamydomonas* and discussed by David M. Bonner and Stanley E. Mills in their book, *Heredity,* part of the Prentice-Hall Foundations of Modern Biology Series.

Biochemical studies of extracts of *Neurospora* have shown that tryptophan, like most other organic molecules of the cell, is synthesized in a sequence of chemical reactions (Fig. 4.1). Mutations in tryptophan-requiring strains map at several genetic locations. Each mutant is defective in one of the steps in the biosynthetic sequence. Those mutations that involve the final reaction, catalyzed by the enzyme tryptophan synthetase (Fig. 4.1), all map in a small section of linkage group 2 called the tryptophan synthetase gene, or *td* gene. Thus, the basic gene-enzyme relationship is clear: a mutation in a specific chromosomal region is reflected in the loss of activity of one enzyme. That the *td* gene

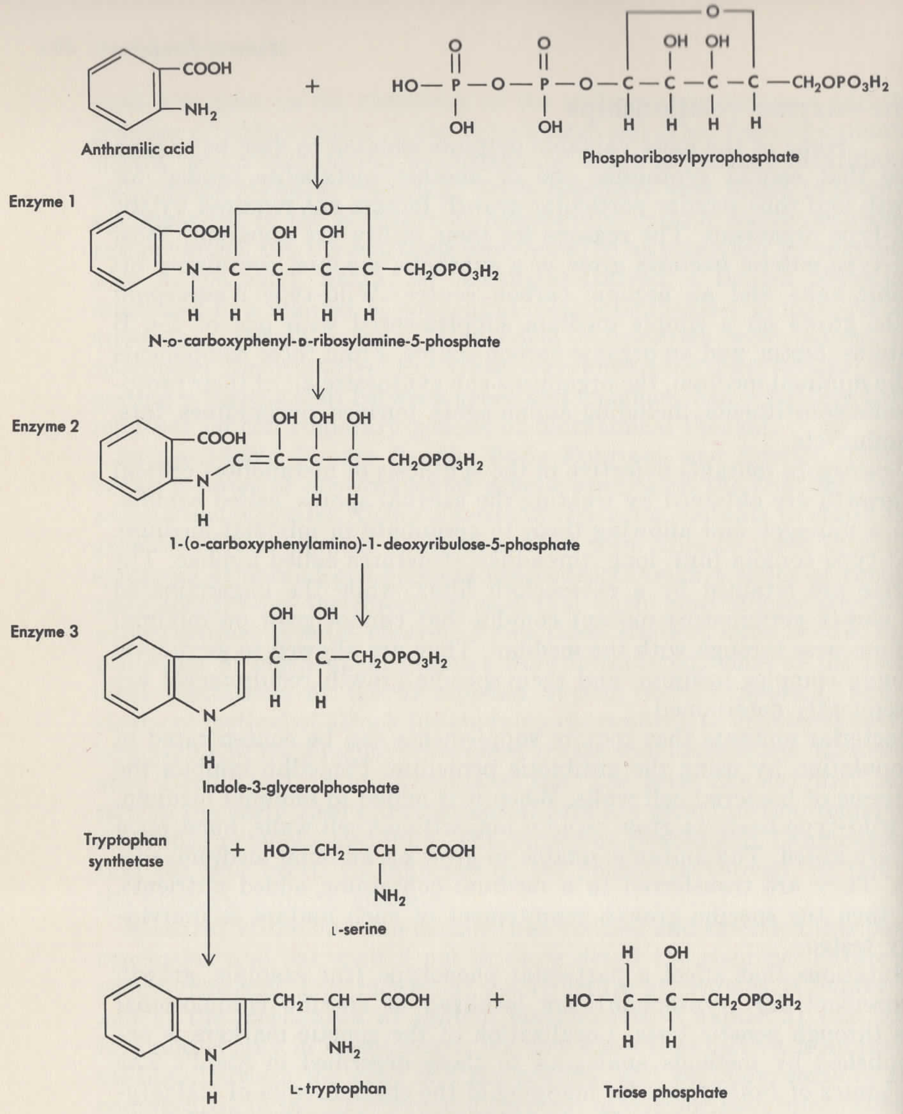

Anthranilic acid
Phosphoribosylpyrophosphate
Enzyme 1
N-o-carboxyphenyl-D-ribosylamine-5-phosphate
Enzyme 2
1-(o-carboxyphenylamino)-1-deoxyribulose-5-phosphate
Enzyme 3
Indole-3-glycerolphosphate
Tryptophan synthetase
L-serine
L-tryptophan
Triose phosphate

Fig. 4.1. Pathway of tryptophan biosynthesis.

is a sequence of adjacent DNA nucleotide pairs that dictate the primary structure of the tryptophan synthetase molecule is now assumed. Further evidence, discussed later in this chapter, is necessary to prove that this assumption is correct.

The genetic block

The reduction in enzymic activity effected by a mutation is termed a genetic block. The block may be complete; that is, it results in complete absence of the enzyme. Or it may be partial (leaky); that is, it results in the formation of altered enzyme that has some limited degree of activity. A particular product, then, is not produced or is produced in insufficient quantity for normal cellular metabolism. Because extreme phenotypic differences usually are easiest to deal with, most biochemical mutants used are those almost totally defective in a particular enzymic activity.

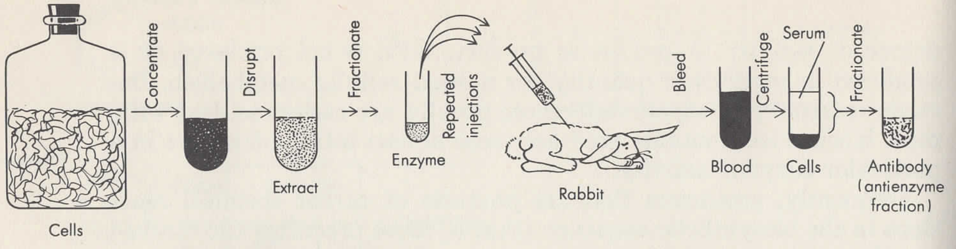
Frequently, precursors that are products of earlier chemical reactions in the biosynthetic sequence (that is, those preceding the blocked reaction) may accumulate in the mutant cells or in the culture filtrate. Identification of some of the precursors accumulated by biochemical mutants has been very helpful in elucidating many pathways of normal intermediary metabolism. The secondary effects of the genetic block are discussed in Chapter 6.

Altered proteins

Genetic crosses between *Neurospora* mutants that cannot form normal tryptophan synthetase demonstrate that many of the mutations are separable by recombination and thus reside at different positions in the polydeoxyribonucleotide chain within the confines of the *td* gene. Some mutants fail to grow at all except in the presence of a tryptophan supplement. In such cases, genetic damage has resulted in the production of an enzyme that is totally unable to catalyze the biosynthesis of tryptophan. Other mutants grow slowly on minimal medium. For example, one mutant, *td24*, forms little or no active tryptophan synthetase at 25° C. and has an absolute requirement for tryptophan at this temperature. However, slow growth on minimal medium ensues when the temperature is elevated to 30° C. This type of mutant is known as a temperature-sensitive mutant. The tryptophan synthetase prepared from *td24* is inhibited by a metal constituent in the cell extract. When the enzyme is separated from the inhibitor by fractionation, activity is restored. The mutant protein is much more sensitive to inhibition by zinc ion, and the energy of activation for tryptophan biosynthesis catalyzed by this mutant protein is much higher than the value obtained with the wild-type enzyme. This in vitro temperature effect is in full agreement with the in vivo temperature effect seen when the mutant is grown at 25° C. versus 30° C. Only at the elevated temperature, where tryptophan synthetase functions at an appreciable rate, is the mutant able to grow, albeit slowly.

Mutant enzymes differ from the wild-type protein and from one another in other ways: electrophoretic mobility, sensitivity to heat or acid denaturation, behavior on column chromatography, and so on.

Perhaps the most striking difference so far found between tryptophan synthetase mutants relates to their ability to form a protein that is inactive enzymatically but has immunological properties similar to tryptophan synthetase. This protein is called cross-reacting material (CRM). CRM is detected by the serological methods outlined in Fig. 4.2 and described in the legend to that figure. Those mutants that produce CRM are called CRM-positive (CRM$^+$) mutants; those that do not are termed CRM-negative (CRM$^-$).


Concentrate
Disrupt
Fractionate
Repeated injections
Cells
Extract
Enzyme
Rabbit
Bleed
Centrifuge
Serum
Fractionate
Blood
Cells
Antibody (antienzyme fraction)

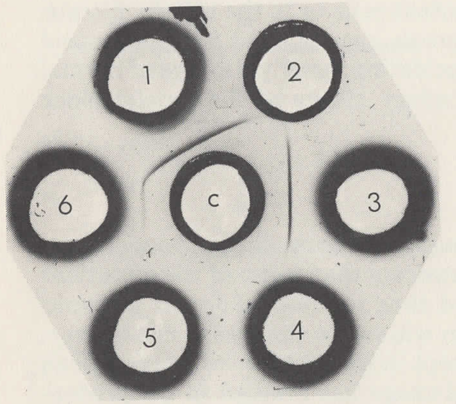
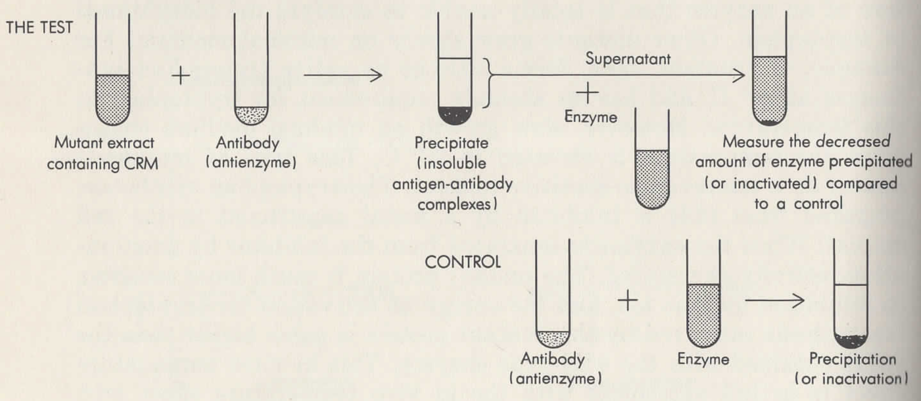


Fig. 4.2. Preparation of immune serum and detection of CRM. (*above*) Some steps in the preparation of immune serum. (*left*) Agar diffusion (Ouchterlony technique). Holes are cut into the agar layer in a petri dish. Antienzyme is added to the center well; enzyme is added to well number 1, and extracts of a CRM$^+$ mutant are added to wells 3 and 6. After diffusion of the antibody molecules and the enzyme protein (and CRM) molecules, a band of precipitation forms between the extract and antibody wells that coalesces with the band formed between the enzyme and antienzyme wells. Extracts of a CRM$^-$ mutant are added to wells 4 and 5, and no band is seen. (*below*) Tube titration (blocking power).


THE TEST
Mutant extract containing CRM
Antibody (antienzyme)
Precipitate (insoluble antigen-antibody complexes)
Supernatant
Enzyme
Measure the *decreased* amount of enzyme precipitated (or inactivated) compared to a control
CONTROL
Antibody (antienzyme)
Enzyme
Precipitation (or inactivation)

Among the CRM$^+$ mutants, immunochemical methods have provided a means of distinguishing several antigenic types of mutationally altered tryptophan synthetase. The formation of these molecules has been related to the genetic locations of mutations within the *td* gene. Furthermore, CRM$^+$ mutants can be classified on the basis of the amount and type of residual enzymatic activity retained by the protein. Wild-type tryptophan synthetase can catalyze at least three biochemical reactions (Fig. 4.3). Some CRM's exhibit no enzymatic activity for any of the reactions. Others retain the capacity to catalyze only one or two of the reactions but lack the ability to catalyze the physiologically important reaction involving the conversion of indoleglycerolphosphate and L-serine to L-tryptophan.

How might CRM$^-$ mutants be explained? The CRM$^-$ phenotype could reflect the formation of unstable molecules, small fragments (free or ribosome-bound), molecules having large regions of the polypeptide chain with an amino acid sequence entirely different from the normal protein, or even polypeptide subunits that have been damaged in certain regions necessary for their association into a specific polymeric form. Indeed, recent evidence suggests that many CRM$^-$ mutants do form a molecule that is a grossly abnormal tryptophan synthetase and fails to give the standard immunological test for CRM.

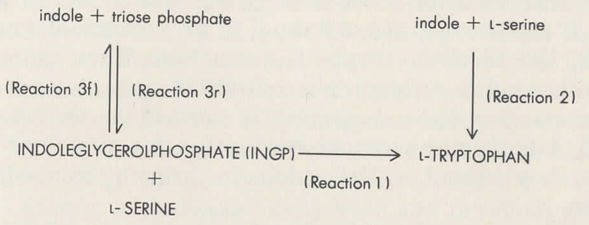

indole + triose phosphate
(Reaction 3f)
(Reaction 3r)
indole + L-serine
(Reaction 2)
INDOLEGLYCEROLPHOSPHATE (INGP)
+
L-SERINE
(Reaction 1)
L-TRYPTOPHAN

Fig. 4.3. Reactions catalyzed by tryptophan synthetase. The normal, physiological reaction is pictured in capitals (reaction 1); accessory reactions are shown in lower case type (reactions 2 and 3). Reaction 3 is separable into two reactions whose interrelationships are not yet clear: a forward reaction (3f) and a reverse reaction (3r) that does not seem to be strictly analogous to a true reversal of reaction 3f.

These results tell us that mutations in a particular gene of *Neurospora* exert their effects at the level of a particular enzyme, tryptophan synthetase, and are reflected in specific changes in the structure and properties of the enzyme. Just as the gene can exist in a variety of forms (alleles), so can the protein. The finding of a variety of altered forms of a protein in mutants for a particular gene constitutes strong evidence that the gene dictates the structure of that particular protein.

More examples of gene-enzyme relationships

In over 100 different gene-enzyme systems that have been studied, a particular gene mutation has been found to result in the loss of a specific enzyme activity. In fewer cases have the investigations proceeded to a point where a gene can be shown to dictate the specificity of the protein, that is, in terms of the production of a defective or mutant protein possessing some but not all of the properties of the wild-type enzyme. However, the number of detailed studies of gene-enzyme systems is being enlarged constantly. In the few examples considered below, some new and accessory principles will emerge and will form a foundation for discussions in later chapters.

The biosynthesis of tryptophan in enteric bacteria proceeds through essentially the same series of reactions as it does in *Neurospora* (Fig.

4.1). The reactions catalyzed by mold and bacterial tryptophan synthetase appear to be similar (Fig. 4.3). However, in contrast to the *Neurospora* enzyme, bacterial tryptophan synthetase is composed of two nonidentical and readily separable protein subunits, termed components A and B. Interactions between the A and B components, which are necessary for maximum catalytic activity, are the subject of a section of Chapter 5. Here we shall consider some general observations on bacterial mutants with defective tryptophan synthetase, that is, those CRM+ mutants that cannot carry out the conversion of indoleglycerolphosphate to tryptophan (Fig. 4.3). Tryptophan synthetase mutants lacking the A component or containing it in an altered form (A-CRM mutants) cannot carry out reactions 1 or 3; but they contain a B component that is normal in all respects and can catalyze reaction 2, the indole-to-tryptophan reaction. Thus, A mutants can grow on either indole or tryptophan. B-CRM mutants contain a defective B component; their A protein is normal, so they can catalyze reaction 3, but they cannot catalyze either reaction 1 or reaction 2. Therefore, they cannot utilize indole for growth, responding only to tryptophan.

In *E. coli,* mutations specifically affecting the A and B proteins are localized in two adjacent sections of the polynucleotide chain known as the *A* and *B* genes, respectively. Recombination tests also show that the *A* and *B* genes are closely linked to two other genes involved in the production of other enzymes in the tryptophan biosynthetic sequence. The clustering of gene loci having related but distinct functions is a subject of Chapter 7.

The top of Fig. 4.4 presents a genetic map of the *A* gene (the lower portion will be referred to in the next section of this chapter). Numbers listed above the line are stock numbers of independently isolated mutants that form A-CRM. As in *Neurospora,* so in *E. coli* tryptophan synthetase mutants, the CRM's exhibit differences in properties. Their differing responses to heat and acid are included in Fig. 4.4. We will return to further examination of A-CRM's after a quick look at some gene-enzyme relationships in the bacterium *Salmonella typhimurium.*

In *Salmonella,* nine genes control the structural properties of nine proteins associated with the biosynthesis of the amino acid L-histidine. As with the tryptophan genes in *E. coli,* the histidine genes are clustered on the chromosome. Fig. 4.5 represents a section of the chromosome containing these histidine genes. Genetic tests indicate that the nine genes are adjacent, or nearly so, and are in the linear order shown in the figure. Several hundred genetically separable sites of mutation have been localized in this region of the genetic map. Each of the mutations solely affects the biosynthesis of histidine. Those mutations mapping within gene *D* always predominantly affect dehydrogenase activity; those in gene *C,* transaminase activity; and so on. Basically, then, mutations in one gene affect one enzyme. The genes are discrete; they do not overlap.

The above description says that one gene controls one enzyme. However, a more precise relationship is the following: *one gene dictates the structure of one polypeptide chain.* There are cases in which a single enzyme or functional protein is composed of more than one type of polypeptide chain. Examples of this type of situation are bacterial tryptophan synthetase and mammalian hemoglobin. Each of these proteins is composed of two types of polypeptide chains, and the structure of each type of polypeptide chain is controlled by a single gene. In the case of bacterial tryptophan synthetase, the two genes are adjacent on the genetic map; in the case of hemoglobin, the genes are not closely linked.

A second reason for relating genes to polypeptide chains, rather than to enzymes per se, is the finding that a single protein may catalyze more than one reaction. One example of such a protein is the *Neurospora* tryptophan synthetase protein already referred to in this chapter (Fig. 4.3). Another protein with more than one activity is L-histidinol dehydrogenase (Fig. 4.5). The dehydrogenase catalyzes two successive reactions: that is, it uses as substrate either histidinol or histidinal. Finally, gene *B* (Fig. 4.5) is credited with dictating the structure of a bifunctional protein with distinct dehydrase and phosphatase catalytic sites. Some single mutational events in gene *B* cause the simultaneous loss of both dehydrase and phosphatase activities. Such a lack is always found in CRM$^-$ mutants. CRM$^+$ mutants lack only dehydrase activity. Furthermore, the phosphatase activity of normal protein may be blocked with the inhibitor arsenate without affecting dehydrase activity.

The general conclusions we want to draw are clear: The gene dictates the structure of a polypeptide chain. In some cases, the single folded polypeptide chain may assume the conformation of a protein with a single (known) enzyme activity. In other cases, several identical polypeptide chains may associate to form an enzyme polymer that catalyzes several reactions. In still other cases, a polypeptide chain of one type may be inactive alone, or have one kind of activity, and gain another activity only when it is associated with a second type of polypeptide chain whose structure is dictated by a second gene.

Alterations in primary structure

In the examples of gene-protein relationships discussed so far, the analyses have shown us only that gene mutations result in changes in the gross structure and properties of a single species of protein molecules. What evidence is there that these reflect changes in the primary structure?

The A protein component of *E. coli* tryptophan synthetase has been purified and subjected to tryptic digestion (see the methods described in Chapter 2). The A protein has a molecular weight of about 29,500 and contains about 282 amino acid residues. The protein is hydrolyzed by trypsin into 25 distinct peptides plus a few additional peptide spots that do not separate well on fingerprinting. Analysis of several of the

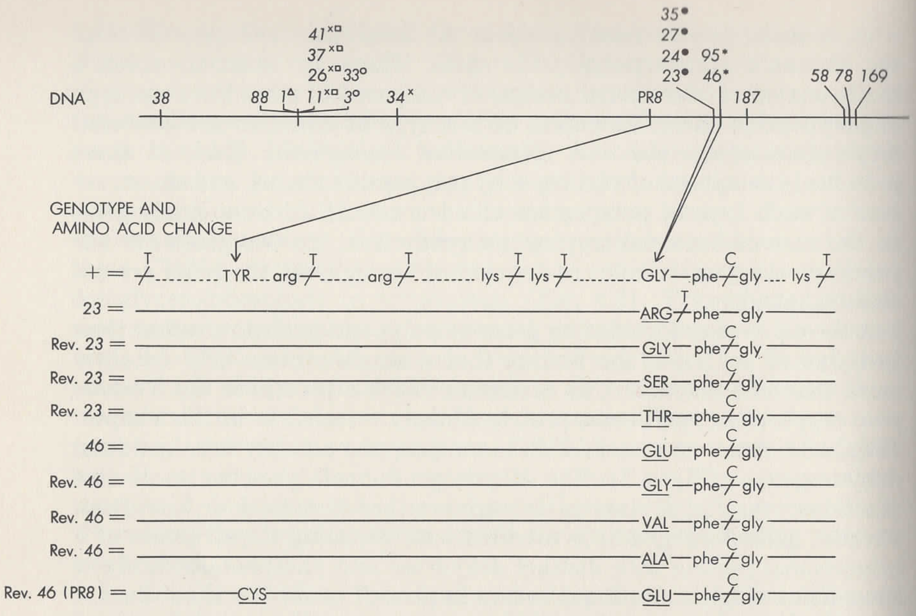

DNA
38
88 1^Δ 11^×□ 3^○
26^×□ 33^○
37^×□
41^×□
34^×
PR8
35●
27●
24● 95*
23● 46*
187
58 78 169
GENOTYPE AND
AMINO ACID CHANGE
+ = TYR arg arg lys lys GLY—phe—gly lys
T C
23 = ARG—phe—gly
Rev. 23 = GLY—phe—gly
Rev. 23 = SER—phe—gly
Rev. 23 = THR—phe—gly
46 = GLU—phe—gly
Rev. 46 = GLY—phe—gly
Rev. 46 = VAL—phe—gly
Rev. 46 = ALA—phe—gly
Rev. 46 (PR8) = CYS GLU—phe—gly

Fig. 4.4. Genetic map of the gene for *Escherichia coli* tryptophan synthetase A protein and some alterations of the wild-type and mutant A proteins. Based on data of C. Yanofsky and collaborators.

The upper portion of the figure shows the genetic map; isolation numbers of some CRM-forming mutants are listed above the line. The numbers in columns indicate that these sites of mutation either are located at identical positions or are extremely closely linked. The most distant mutations give between 800 and 1,600 times the frequency of wild-type recombination found between the closest two mutations (e.g., *A23* and *A46*); this value may be compared with the content in the A protein of about 280 amino acids.

Some of the properties of the mutant proteins are noted as superscripts to the mutant numbers. They indicate that the respective CRM's are: heat labile (●); heat labile, but more stable than *A23* (Δ); heat sensitive to the same degree as the wild-type protein (×); heat stable (O); heat stable, but less stable than *A3* or *A33* (*); acid precipitable (□).

The first line in the lower portion of the figure depicts the amino acid sequence in a portion of the wild-type A protein. T and C indicate points of hydrolysis by tryspin and by chymotrypsin, respectively. Dots represent other amino acids. Beneath the wild-type A protein are A-CRM'S in which amino acid replacements have been detected at the positions underlined. Some of the proteins are from revertants (Rev.) that have regained full or partial tryptophan synthetase activity.

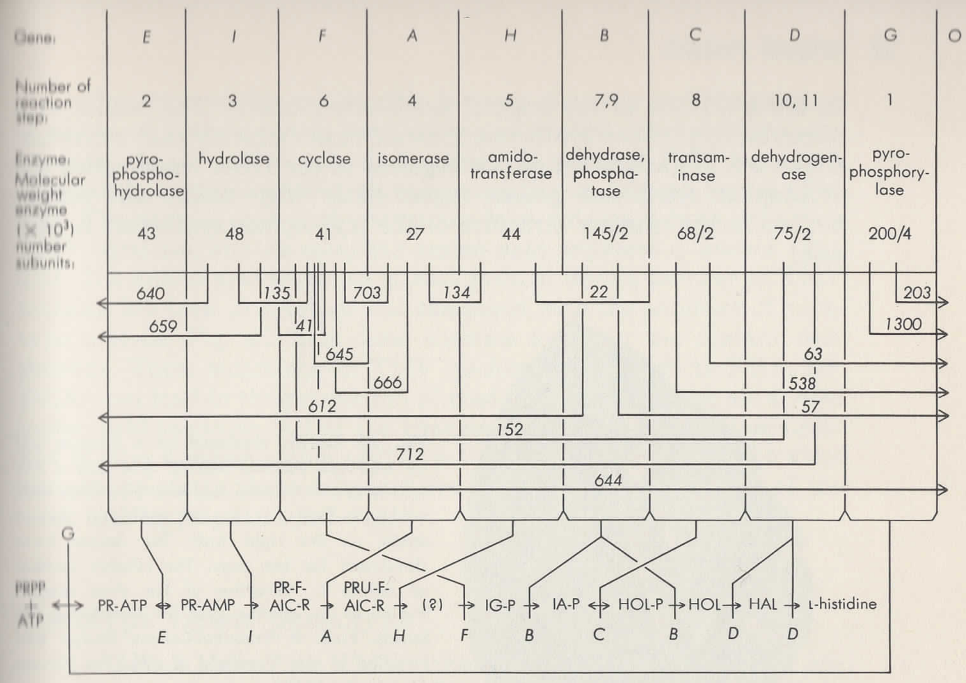

Gene:
Number of reaction step:
Enzyme:
Molecular weight enzyme (× 10³) number subunits:
E
2
pyro-phospho-hydrolase
43
I
3
hydrolase
48
F
6
cyclase
41
A
4
isomerase
27
H
5
amido-transferase
44
B
7,9
dehydrase, phospha-tase
145/2
C
8
transam-inase
68/2
D
10, 11
dehydrogen-ase
75/2
G
1
pyro-phosphory-lase
200/4
O
640
659
135
41
645
703
666
612
134
22
152
712
644
203
1300
63
538
57
G
PRPP + ATP
PR-ATP
PR-AMP
PR-F-AIC-R
PRU-F-AIC-R
(?)
IG-P
IA-P
HOL-P
HOL
HAL
L-histidine
E
I
A
H
F
B
C
B
D
D

Fig. 4.5. Genetic structure of the histidine region of the *Salmonella* chromosome and the reactions catalyzed by the gene products. Nine genes, *A-I*, specify the structures of nine proteins involved in the biosynthesis of L-histidine. Because two of the enzymes are bifunctional, the reaction pathway consists of eleven steps. The molecular weight of each enzyme is indicated along with the number of polypeptide chain subunits that comprise the enzyme, if this is known to be greater than one. The numbers below the line are stock numbers of multisite or extended mutations, and the bars indicate the extent of the genetic lesions. The following abbreviations are used: adenosine triphosphate (ATP), phosphoribosylpyrophosphate (PRPP), phosphoribosyl adenosine triphosphate (PR-ATP), phosphoribosyl adenosine monophosphate (PR-AMP), phosphoribosyl formamidino aminoimidazole carboxamide ribotide (PR-F-AIC-R), phosphoribulosyl formamidino aminoimidazole carboxamide ribotide (PRU-F-AIC-R), an unidentified compound (?), imidazoleglycerol phosphate ester (IG-P), imidazoleacetol phosphate ester (IA-P), L-histidinol phosphate ester (HOL-P), L-histidinol (HOL), and L-histidinal (HAL).

tryptic and chymotryptic peptides of the mutant proteins has revealed the presence of single amino acid substitutions in some of them. That is, a particular amino acid that is found in a specific peptide in a fingerprint profile of the wild-type enzyme has been replaced by a different amino acid in the mutant enzyme. The lower portion of Fig. 4.4 shows several amino acid substitutions.

Most revealing is the analysis that has been made of peptides from A-CRM's of tryptophan synthetase mutants *A23* and *A46*. Fingerprints of tryptic peptides from *A23* CRM showed one wild-type peptide to be missing; this peptide was counterbalanced by the appearance of two new peptides not found on fingerprints of wild-type protein. The mutant protein was assumed to contain an extra trypsin-sensitive residue (arginine or lysine) that is not present in the wild-type protein. An arginine residue is shown in the diagram of Fig. 4.4, because, upon further analysis, arginine actually proved to be the amino acid that was substituted for glycine at that position in the wild-type protein. Through further carefully controlled enzyme digestions and amino acid analyses, the alteration in the *A46* CRM was shown to involve replacement of this same glycine residue by a glutamic acid residue. As indicated on the genetic map, the *A23* mutant site is very close to, but nevertheless distinct from, the *A46* site. This is strong evidence that more than one nucleotide pair is responsible for determining the specific location of a particular amino acid in a protein.

Finally, a pair of double mutants, *A23*, *A33* and *A11*, *A46*, was obtained in recombination experiments. Each of the double mutants had two alterations in the *A* gene. The A-CRM's of each of the double mutants were found to contain both of the peptide alterations noted in the A-CRM's of the original single-site mutants. The results with bacterial tryptophan synthetase indicate most elegantly that individual single-site mutations within the segment of the DNA chain called the tryptophan synthetase genetic region cause single amino acid substitutions in the primary structure of the tryptophan synthetase A protein.

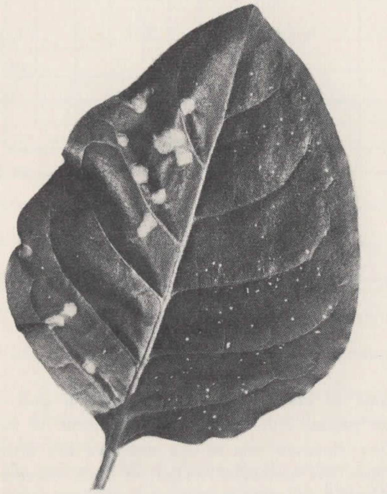


Fig. 4.6. Lesions produced on a tobacco leaf by wild-type tobacco mosaic virus and a mutant strain. A tobacco leaf was inoculated with wild-type TMV on the left half and mutant strain on the right half. The lesions have developed for ten days. The relative number of lesions is indicative of the virus concentration in the solution used for inoculating the leaves. From H. Fraenkel-Conrat, *Design and Function at the Threshold of Life: The Viruses* (Academic, 1962).

Fig. 4.7. Amino acid substitutions found in various mutant strains of tobacco mosaic virus. The complete amino acid sequence of wild-type TMV protein is shown in Fig. 2.4. Mutants were obtained by isolating, from lesions on tobacco leaves inoculated with nitrous acid–treated virus, mutant strains with abnormal growth properties. In 41 mutant strains that showed amino acid substitutions, the substitutions occurred at 22 different positions in the protein. Some of the remaining substitutions were found to be identical to others, indicating that the same mutational event may have occurred more than once. Some examples of the substitutions are shown here. Based on data of H. G. Wittmann.

Residue number	Amino acid present in wild-type protein	Amino acid present in mutant protein
5	thr	ileu
20	pro	leu
24	ileu	val
55	ser	leu
59	thr	ileu
61	arg	gly
63	pro	ser
66	asp	gly
73	asp-N	ser
97	glu	gly
107	thr	met
126	ileu	val
134	arg	gly
138	ser	phe
156	pro	leu

Confirmation of this relationship between the gene and the primary structure of protein comes from studies of several other systems—for example, the tobacco mosaic virus protein. Mutants of TMV have been isolated that exhibit altered growth properties on tobacco leaves (Fig. 4.6). The protein coat of the virus has been extracted from wild-type TMV particles, and its complete amino acid sequence is known (Fig. 2.4). The amino acid sequence of coat protein from a number of TMV mutants has been determined and compared with the sequence of wild-type protein. Fig. 4.7 lists some changes found in the mutant coat proteins. Thus, mutations in TMV nucleic acid (which is RNA, not DNA) can lead to the production of coat protein possessing an altered amino acid sequence. All of the changes are amino acid replacements and involve a single amino acid at a time. In these respects, the virus mutants reinforce the conclusions reached from our discussion of the work with tryptophan synthetase.

Genotype versus phenotype

About half of the TMV mutants that have been isolated appear to have normal coat protein. Possibly, in these, mutation affects some function other than the production of the coat protein, for the virus produces other proteins during its intracellular growth that are not incorporated into the mature particle; however, some amino acid changes in proteins are not easy to detect. For example, the tryptic and chymotryptic peptides of A-CRM's from *E. coli* tryptophan synthetase mutants *A3*, *A34*, and *A75* are indistinguishable from the peptides of the wild-type protein on fingerprint analysis. Here it is quite possible that an amino acid substitution has occurred in one of the peptides but that the substitution involves an uncharged or neutral amino acid, so that the movement of a peptide during chromatographic and electrophoretic separation procedures is not altered. In such cases, a more complete analysis is essential to detect any alteration that may be present but unresolved by this technique.

Generally, however, the fingerprinting technique does give evidence of amino acid changes and, sometimes, detects changes that are unsuspected. For instance, mutant *A23* occasionally mutates to strains that are again able to grow in the absence of tryptophan and exhibit considerable tryptophan synthetase activity. The A proteins of some of these revertants again contain the glycine originally present in the wild-type A protein; other revertants contain threonine or serine, however (Fig. 4.4). Those with threonine grow slowly in the absence of tryptophan, but those with serine cannot be distinguished phenotypically from wild type. Functionally, the serine-containing proteins are normal. This observation is quite important, for it proves that some mutations in the genetic material normally can go undetected. If the amino acid replacement occurs in a noncritical portion of the molecule and involves the replacement of one amino acid by another possessing similar size and properties, the protein may still retain wild-type activity although altered in its primary structure.

A similar situation is found among revertants of mutant *A46* (Fig. 4.4). Some revertants are true wild types (resubstitution of GLY), some mimic wild type (substitution of VAL), and others are partially effective in the absence of tryptophan (substitution of ALA). In still another class of revertants, the mutant tryptophan synthetase still contains the original amino acid substitution (glutamic acid in place of glycine); in addition, it contains a second amino acid substitution, a cysteine for a tyrosine. These revertants, which can grow slowly in the absence of tryptophan, are double mutants. Each of the single mutations leads to inactive proteins with single amino acid substitutions; together the two mutations lead to protein with two amino acid substitutions and partial activity. In some manner, the second substitution modifies the configuration of the protein molecule so that it is capable of functioning as a catalyst, although its catalytic efficiency is less than that of the normal protein.

This secondary mutation is termed a suppressor mutation because it suppresses the expression of the original mutant phenotype (tryptophan requirement), although the original primary mutation remains present in unaltered form. And it is an intragenic suppressor mutation because it is located in the same gene as the primary mutation.

One can see, from these examples, that each change in the genotype is reflected in a change in an amino acid in a specific protein. Even more, they emphasize that the effect of the amino acid substitution on the activity and other properties of the protein is dependent both on the specific amino acid residues involved and on the location in the polypeptide chain.

Gene mutations and protein structure: Some generalities

The gene is pictured as a sequence of adjacent nucleotides in the polymer deoxyribonucleic acid. The gene is divisible by mutation, because any single nucleotide can be inserted in place of another. The gene is divisible by recombination, for recombinations can take place between pairs of adjacent nucleotides along the polynucleotide chain. The gene is basically a functional unit. Its function lies in the control of the structure of a single class of polypeptide chains.

The amino acid sequence of a polypeptide chain is dictated by the sequence of nucleotides in the corresponding gene. Mutations at different positions within the gene ultimately result in alterations in the structure of the polypeptide chain. A single-site mutation causes the substitution of one amino acid by another in the polypeptide chain. Thus, we can conclude that certain genes control the primary structure of individual and specific protein molecules, and that mutations in a structural gene result in changes in the primary structure of the protein molecule. The amino acid replacements, giving rise to mutant proteins, are qualitative changes. The particular location in which the amino acid substitution occurs within the polypeptide chain is important in determining the ultimate effect of the mutation on the biological activity and other properties of the protein molecule. Also, the type of amino acid substitution is important. Substitution of one amino acid by another containing a very dissimilar side chain generally has a more extreme effect than does replacement with more similar amino acids. In fact, where substitutions involve very similar amino acids, the effect on the configuration and, therefore, on the properties of the protein may be very slight. Such alterations in structure may be detected only by detailed chemical analysis of the purified protein.

The polypeptide chain is a linear structure (Chapter **2**). This also seems to be true for the mRNA (Chapter 3) and for the DNA. In fact, all current results strongly suggest that the three classes of macromolecules are colinear, as predicted by the model seen in Fig. **3.12**. In effect, the concept of colinearity says that mutations that are distant from one another in a gene will cause amino acid changes in widely different locations of the polypeptide chain. For example, mutations at the extreme opposite ends of a gene should determine the nature of the N-terminal and the C-terminal amino acid residues, respectively. Mutations at the same position, or very close together on the genetic map, would be expected to cause amino acid substitutions at the same location, or at adjacent locations, in the polypeptide chain. As we indicated in Chapter 1, the concept of colinearity now has received firm experimental verification.

References

Cushing, J. E., and D. H. Campbell, *Principles of Immunology*. New York: McGraw-Hill Book Company, 1957. An introductory text to immunology and its applications in biology.

Ingram, V. M., "The Evolution of a Protein," *Federation Proc.*, *21* (1962), 1053. A short discussion of changes in the primary structures of proteins, including the original change discovered by the author, and their potential significance to the field of evolution.

Yanofsky, C., "The Tryptophan Synthetase System," *Bact. Rev.*, *24* (1960), 221. A review of experimental observations on *Escherichia coli* and *Neurospora* tryptophan synthetases.

———, and P. St. Lawrence, "Gene Action," *Ann. Rev. Microbiol.*, *14* (1960), 311. One of the better modern reviews of advances in biochemical genetics.

Questions

4.1. What is meant by the term *genetic block?* Is a genetic block always complete? Why?

4.2. Look up the original literature and review in about 300 words plus a one-page diagram an additional case of a gene-protein relationship. You may find an appealing example on your own, or you may wish to select your example from among:
(*a*) Lysozyme of T4 bacteriophage.
(*b*) Utilization of a sugar or biosynthesis of a purine, pyrimidine, or amino acid in *Escherichia coli, Salmonella typhimurium, Neurospora crassa,* or *Saccharomyces cerivisiae.*
(*c*) Xanthine dehydrogenase, amylase, or esterases of *Drosophila melanogaster.*
(*d*) Glucose-6-phosphate dehydrogenase in man.

4.3. Normal adult hemoglobin, hemoglobin A, consists of four polypeptide chains, two alpha chains and two beta chains. The primary structures of the two kinds of chains have some similarities but differ at many positions.
(*a*) Is the structure of hemoglobin A dictated by one or by two genes? Justify your answer.
(*b*) A difference in electrophoretic mobility (surface charge) was detected between normal and abnormal hemoglobin by Linus Pauling and coworkers in 1949. What structural alteration in the protein might underlie such a change?
(*c*) Some of the electrophoretically abnormal hemoglobins lead to diseased states (e.g., sickle cell anemia; see Fig. 6.7), while other abnormal hemoglobins are without effect on the phenotype. Explain why such different responses might be predicted.

4.4. What tests does a geneticist apply to determine if two independently isolated mutations affect the same or different genes?

4.5. In tabular form, set out the probable relationships between chemical terms and genetical terms:

	Series of adjacent nucleotide pairs dictating structure of polypeptide chain	Smallest unit of mutation	Smallest unit indivisible by recombination	Smallest unit in which recombination occurs
Chemical term (in relation to structure of DNA)				
Genetical terms that apply				

Five

Complementation

In the last chapter, the function of the gene was described in terms of the genetic control of the primary structure of polypeptide chains. The sequence of bases in the polynucleotide chain dictates the positioning of amino acid residues at specific sites in the polypeptide chain. The evidence is clear-cut that gene mutations can cause replacements of amino acids in protein. The analysis of physical and chemical changes in the structure of a protein molecule requires that the protein be pure, and comparatively few proteins in genetically defined systems have been obtained in the quantities and state of purity required for such analyses. At the present time, relatively few genetically controlled activities in higher organisms or even in microorganisms have been described in general biochemical terms, in enzymological terms, and finally in terms of amino acid substitutions within a specific protein. Clearly what we need is an additional, more rapid method to determine whether two different mutations affect the same or separate genic functions.

The average chemical length of a single gene is estimated to be one or more thousand nucleotide pairs. Recombination tests can tell us if the sites of two mutations are closely linked and are contained in the relatively short polynucleotide sequence comprising a single gene. If the two mutations lie on different chromosomes, on different arms of the same chromosome, or far apart on a single chromosome arm, they are usually assumed to affect the structure of two different polypeptide chains. Thus, recombination tests provide us with an initial screening method for determining whether two mutations affect the same or different functions. Where the two mutations are close together on the chromosome, the question of whether the same or different polypeptide chains are involved must be answered by additional experimental methods.

Getting genes together

An independent genetic test, the *cis-trans* test, can give us much more information about the functional identity of or difference between two mutations. The *cis-trans* test is an in vivo method of evaluating the phenotypic effect of two independently occurring mutations by having both mutations present in several alternative configurations in a single cell. The following brief discussion should make clear the difference between the recombination method of analysis and the *cis-trans* test, or complementation test, as it is frequently called.

Two independently isolated mutations, *a* and *b*, are found to be closely linked and thus could possibly be located within the same gene. They may be brought together on the same chromosome through genetic recombination. The recombinants will be wild type (++) and the double mutant (*ab*). Along with the two parental types, (*a*+) and (+*b*), we then have four marker configurations possible among the progeny. If we stipulate that we want the two different mutations in the same cell, there are two ways this may be achieved. The two mutations may be carried by one chromosome (the double mutant *ab*), while the wild-type alleles are on the second chromosome (++). This is the *cis* configuration. The alternative configuration occurs when the two mutations are located on different chromosomes as denoted by *a*+ and +*b*. This is called the *trans* configuration.

The *cis-trans* test does not depend upon genetic recombination; it depends merely upon genetic function. The main technical problem in the test is the placement of two different chromosomes, or critical portions of chromosomes, together in the same cell. The manner in which this is achieved varies with different organisms, their modes of reproduction, and their normal genetic complement. Figure 5.1 shows some of the primary methods used at the current time to execute the *cis-trans* test. The figure illustrates only the operations involved in obtaining the important *trans* configuration.

Fig. 5.1. General methods for bringing together two mutations (in the *trans* configuration) in single cells.

(a) *Heterozygous diploids* formed after meiosis and fertilization.

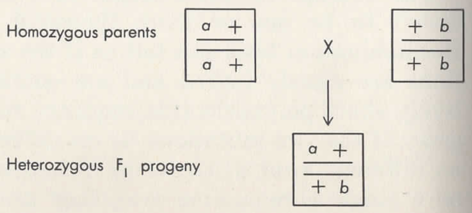
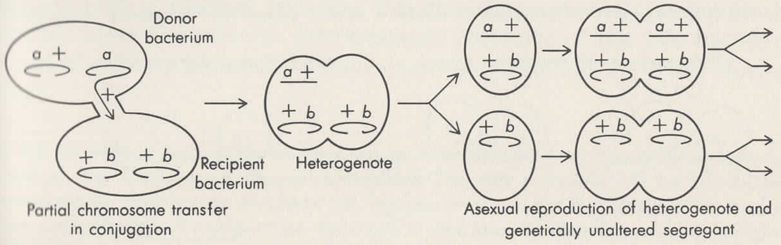

Homozygous parents
a +
a +
X
+ b
+ b
Heterozygous F₁ progeny
a +
+ b

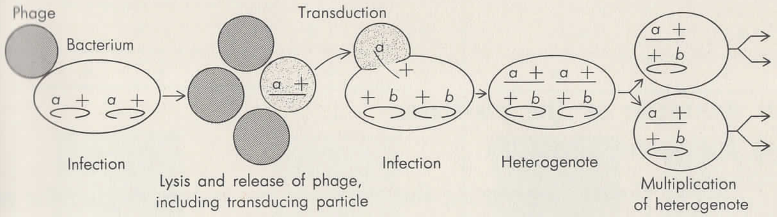
Fig. 5.1. (cont.)

(b) *Partial diploids in bacteria:*

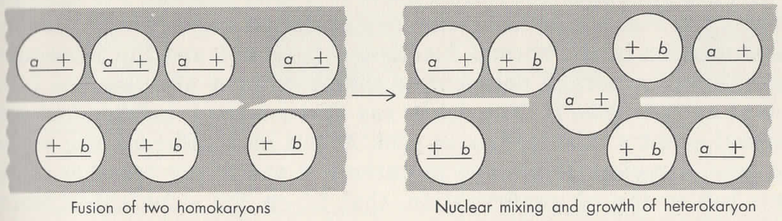
(1) *Following conjugation.* A portion of the chromosome is transferred from the donor cell to the recipient cell. Immediately after gene transfer, the effects of the two elements together in common cytoplasm may be studied. In special cases, the transferred chromosome fragment remains in diploid condition (heterogenote) during subsequent divisions of the recipient bacterium.


Donor bacterium
a +
a
+
+ b
+ b
Recipient bacterium
a +
+ b
+ b
Heterogenote
a +
+ b
a +
a +
+ b
+ b
+ b
+ b
+ b
Partial chromosome transfer in conjugation
Asexual reproduction of heterogenote and genetically unaltered segregant

(2) *Following transduction.* When certain bacterial viruses are released from infected cells, they occasionally pick up short chromosome fragments from the killed "donor" bacteria. These chromosome fragments may enter other (recipient) bacteria. In some situations, the fragment multiplies in synchrony with multiplication of the host chromosome (as in the heterogenote shown here). On other occasions, the transduced fragment does not multiply following penetration into the recipient bacterium; it merely is preserved as a single genetic element and is able to function. In this case (called abortive transduction), at each succeeding division of the recipient bacterium, one cell is a genetically altered partial diploid and contains the transduced fragment, while the other daughter cell lacks the abortively transduced fragment.


Phage
Bacterium
a + a +
Infection
a +
Lysis and release of phage, including transducing particle
Transduction
a
+
+ b + b
Infection
a + a +
+ b + b
Heterogenote
a +
+ b
a +
+ b
Multiplication of heterogenote

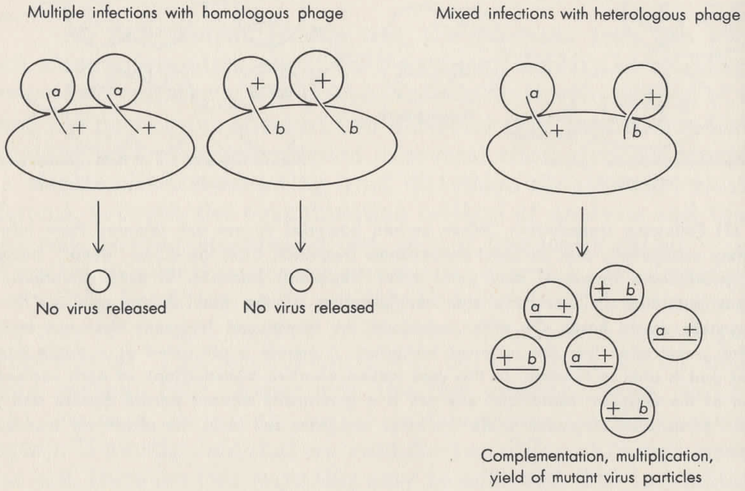
(c) *Heterokaryon* formed by cell fusion and presence of genetically different haploid nuclei in common cytoplasm. No nuclear fusion (diploidy) occurs.


a +
a +
a +
a +
+ b
+ b
+ b
Fusion of two homokaryons
a +
+ b
+ b
a +
a +
+ b
+ b
a +
Nuclear mixing and growth of heterokaryon

(d) *Heterokaryon containing disomic nuclei.* Occasionally, spores are found that carry an extra chromosome (that is, $n + 1$). When such a spore germinates, some of the nuclei contain an extra chromosome, but others lose the extra chromosome. If the spore had originated from a genetic cross between strains *a*+ and +*b*, a heterokaryon (pseudo-wild type) may be formed.

Fig. 5.1. (*cont.*).

(e) *Mixed infection by viruses.* This test is made most conveniently under conditions where neither of two genetically different virus particles alone can successfully infect the cell. If the two mutations are in separate genes, the individual virus particles can replicate together in a common cytoplasm. Mutant virus particles, genetically identical with each of the two kinds of the parental infecting particles, are released after the infection has terminated. No genetic recombination is necessary for production of virus progeny through complementation, although the released virus may include a normal proportion of genetic recombinants.

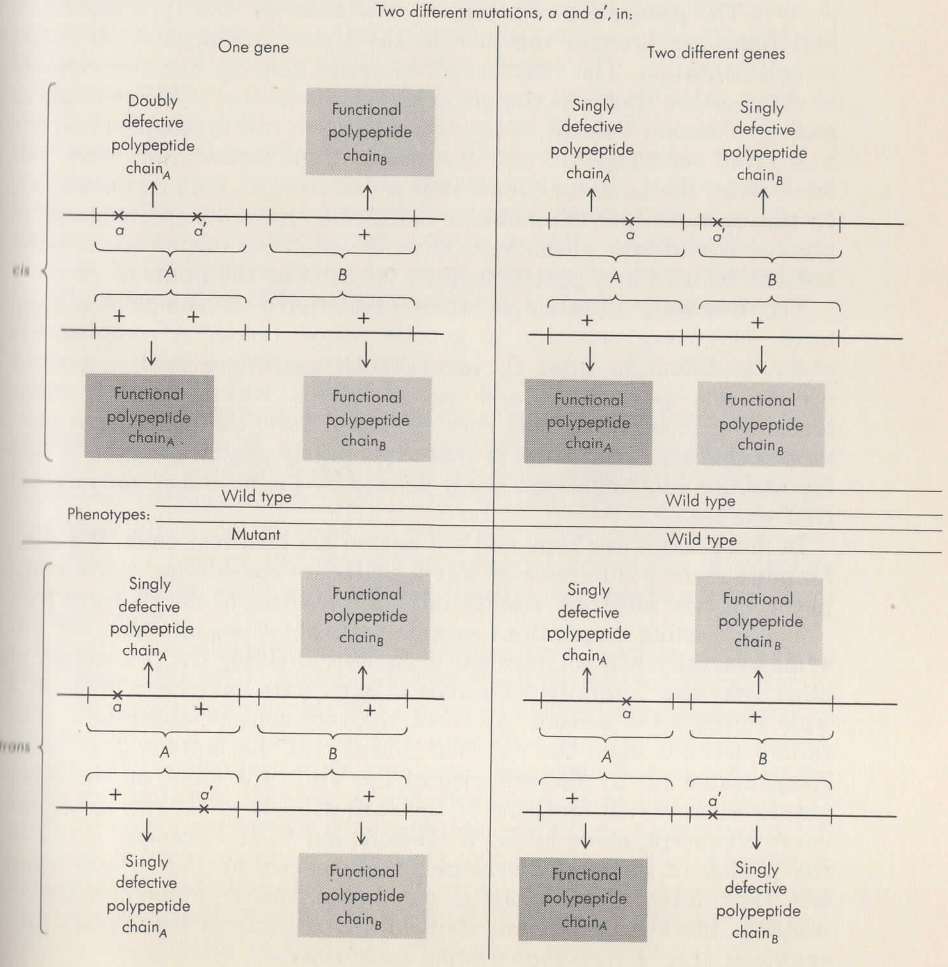

Multiple infections with homologous phage
Mixed infections with heterologous phage
a
a
+
+
+
+
b
b
a
+
+
b
No virus released
No virus released
a +
+ b
a +
+ b
a +
+ b
Complementation, multiplication, yield of mutant virus particles

An example of the *cis-trans* test for gene function

The theory underlying the *cis-trans* test is based upon the one gene, one polypeptide chain concept. In thinking about the *cis-trans* test, it may be helpful to consider that genes act as discrete elements that function independently of one another, although some qualification of this general statement will be made in a later chapter. For example, referring to Fig. 5.2, let us assume that mutation *a* specifically prevents the expression of the normal function of a gene *A* and the subsequent formation of a normal polypeptide chain$_A$ without affecting the function of another gene *B*. Now let us say that a second, independent mutational event *a′* also occurs in gene *A*, but at a different site within the gene *A* region. If the two mutations, *a* and *a′* are together in the same cell in the *cis* configuration, that is, on the same chromosome, there will be produced *X* quantity of a doubly defective polypeptide$_A$, *X* quantity of wild-type polypeptide$_A$, and 2*X* quantity of wild-type polypeptide$_B$ (upper left in Fig. 5.2). Since both normal polypeptide$_A$ and normal polypeptide$_B$ are present, the phenotype will approach that of wild type, and, in fact, may be quite similar to it.

Now examine this situation with regard to the same two genetically separable mutations, *a* and *a′*, when they are in the *trans* configuration, that is, when they are present on different chromosomes (lower left in Fig. 5.2). Each of the *a*+ and the +*a′* genotypes causes the formation of a defective polypeptide$_A$. No functional polypeptide$_A$ is produced, and the phenotype of the cell will be mutant even though functional polypeptide$_B$ is formed. Whether the phenotype is wild type or mutant depends upon the relative chromosomal positions of the two mutations *a* and *a′*. This phenomenon sometimes is called the *cis-trans* position effect.

Fig. 5.2. Theoretical basis of the *cis-trans* test for functional allelism. Two different mutations, *a* and *a′*, are placed in a common protoplasm. The two mutations can be on the same chromosome or chromosome fragment (*cis* configuration) or on two different chromosomes (*trans* configuration). The phenotype discerned in the *trans* configuration affords insight to whether two mutations are in the same gene or in two separate genes.


Two different mutations, a and a′, in:
One gene
Two different genes
cis
Doubly defective polypeptide chain$_A$
Functional polypeptide chain$_B$
Singly defective polypeptide chain$_A$
Singly defective polypeptide chain$_B$
a
a′
+
A
B
+
+
+
a
a′
A
B
+
+
Functional polypeptide chain$_A$
Functional polypeptide chain$_B$
Functional polypeptide chain$_A$
Functional polypeptide chain$_B$
Phenotypes:
Wild type
Mutant
Wild type
Wild type
trans
Singly defective polypeptide chain$_A$
Functional polypeptide chain$_B$
Singly defective polypeptide chain$_A$
Functional polypeptide chain$_B$
a
+
+
A
B
+
a′
+
a
+
A
B
+
a′
Singly defective polypeptide chain$_A$
Functional polypeptide chain$_B$
Functional polypeptide chain$_A$
Singly defective polypeptide chain$_B$

Examination of the right side of Fig. 5.2 will convince you that a single dose of functional polypeptide$_A$ and a single dose of functional polypeptide$_B$ will be produced in both the *cis* and the *trans* configurations if the sites of the mutations a and a' are in separate genes. Note that it is the *trans* test that critically distinguishes the alternative possibilities. If the two mutations are in the same gene, the *trans* phenotype is mutant; if the two mutations are in different genes, the *trans* phenotype is wild type. In contrast, the phenotype of the *cis* configuration is always wild type. Thus, the *cis* test is valuable only as a control. The appearance of the wild-type phenotype in the *cis* test informs us that the mutations are recessive. That is, the wild-type activities continue to be expressed even in the presence of the mutant genes on the opposite genetic element.

One gene, one function

The production of a phenotype similar to wild type when two mutations are brought together in the *trans* configuration is called complementation. The function of an active gene on one chromosome or chromosome fragment complements the function of a different active gene on a second element. When two mutations fail to complement, the wild-type phenotype is not discerned. Complementation does not occur when the same functional unit is defective on both chromosomes. In this case, neither chromosome can fulfill the functions required to produce a wild-type phenotype. Thus, the *cis-trans* (complementation) test can be used as a genetic method for defining the limits of genes.

The following situation is often encountered in complementation tests: (1) every mutation in genetic region (gene) *A* complements every mutation in gene *B*, and (2) no mutations within gene *A* complement each other and no mutations within gene *B* complement each other. In this very restricted case, the genes have been termed cistrons. The cistron is defined by purely genetical methods, and represents a functional unit solely defined by the results of unequivocal *cis-trans* tests.

In many cases, the gene and the cistron are identical units. We shall examine several situations of this type before considering a few complications that affect the results and interpretation of the *cis-trans* test.

The preceding descriptions assume that mutations affecting one enzyme generally will complement mutations affecting the production of other enzymes. Literally thousands of such tests have been made in a wide variety of organisms. After all, the very concept of recessive mutation depends upon the deficient activity of the mutant gene being compensated for by the presence of the wild-type allele on the other genetic element! In fact, one of the earliest tests of the one gene, one enzyme concept, made by G. W. Beadle and V. L. Conroodt, involved the testing of different biochemical deficiencies by the heterokaryon test (Fig. 5.1c). These workers noted that two tryptophan-deficient mutants, blocked at different steps in the pathway of tryptophan biosynthesis (Fig. 4.1), complemented each other.

A representative example of complementation behavior between mutants altered in different enzymes is diagrammed for the histidine biosynthetic pathway (Fig. 4.5). Mutations in one of the genes *A* through *I* complement mutations located in any of the remaining genes. Together in the *trans* configuration, pairs of mutations from different genes usually give a phenotype identical with or approximating wild type. Over 5,000 such pair-wise combinations have been tested.

Some exceptional mutants fail to complement strains that are mutant for different genes. Such exceptional mutants are found, by recombination tests, to have gross changes or aberrations that involve a segment of the polynucleotide chain that extends into two or more adjacent genes. The extent of some of these multisite mutations, measured by recombination tests, is indicated by the numbered brackets under the chromosome in Fig. 4.5. The complementation behavior is accounted for by the position and extent of the genetic change. For example, histidine mutant *703* complements mutants of genes *E*, *I*, *H*, *B*, *C*, *D*, and *G*, but does not complement with mutants of either gene *F* or gene *A*. Recombination tests show that mutant *his703* lacks the "right" end portion of gene *F* and the "left" portion of gene *A*.

Several hundred complementation tests between different pairs of mutants located within each of several genes *F*, *A*, *H*, *C*, and *G* show that each gene is a separate cistron. That is, all possible pairs of *F* mutants that were tested for complementation with each other failed to do so. Thus, gene *F* is a single functional genetic unit, gene *A* is a separate functional genetic unit, and so on. The *cis-trans* test is extremely useful in these particular cases, because the genes with related functions are adjacent to each other on the bacterial chromosome.

One gene, one polypeptide chain

In the description immediately preceding, one functional unit or cistron (defined by the *cis-trans* test) corresponded to one enzymic activity. This is often found. However, some functional proteins require two cistrons. Close examination of such systems demonstrates that these functional proteins are composed of two polypeptide chains. One gene, or cistron, dictates the structure of one of the polypeptide chains, while the other gene dictates the structure of the second polypeptide chain.

Recall, from Chapter 4, the reactions carried out by tryptophan synthetase (Fig. 4.3). Also recall that the bacterial enzyme active in the physiologically significant reaction, indoleglycerolphosphate to tryptophan, is a composite of two dissimilar proteins, components A and B, each controlled by its own gene. Complementation tests in the *trans* configuration between A and B mutants demonstrate that (1) all A mutants complement all B mutants, and (2) no complementation is found between pairs of A mutants or between pairs of B mutants. The *cis-trans* test thus unequivocally demonstrates the presence of two independent genes, or cistrons, involved in the production of the two protein components of bacterial tryptophan synthetase.

Several other attributes of the interactions of the two components of tryptophan synthetase are of interest in physiological genetics. Recall that the B protein can by itself catalyze the indole-to-tryptophan reaction. All mutants defective in gene *A*, responsible for A protein, can grow on indole, because each mutant possesses a B protein. Thus, this residual accessory reaction, catalyzed by the B protein, becames physiologically important to bacteria that are mutant in the *A* gene. It allows them to grow on indole as an alternate growth factor to tryptophan. If tryptophan is lacking from the environment, they can still multiply at the expense of any indole that may be present.

The bacterial tryptophan synthetase A mutants are divided into two classes: those containing an altered protein immunologically closely related to active A protein (CRM$^+$ mutants) and those containing no detectable A-CRM (CRM$^-$ mutants). These two classes of mutants can be distinguished from each other on another basis, namely, by their rates of growth in the presence of indole. The reason for this difference becomes apparent when the indole-to-tryptophan reaction is studied with purified B protein in vitro. While B protein can catalyze this reaction without A protein present, it does so at only 10 per cent of the rate achieved in the presence of A protein. It appears that combination of B protein with A protein elicits some conformational change in B protein that enormously enhances its catalytic activity. Similarly, A protein alone can catalyze the indoleglycerolphosphate-to-indole reaction in the absence of the B protein, yet in the presence of B protein the reaction rate is tremendously increased. Consequently, it would appear that both A and B protein function at peak efficiency only when in a binary complex with each other.

The results collectively demonstrate that protein-protein interactions can play important roles in physiological regulation of the rates of enzymic reactions. The results show conclusively that the gene, or cistron, defined by genetic tests, is involved with the production of a single polypeptide chain. It is important to keep in mind the fact that complementation is due to interaction of gene products, not of the genes themselves.

Intragenic complementation

Very often in biology, exceptional cases are found to seemingly clear-cut rules. When investigated thoroughly, these exceptions frequently furnish important contributions to our basic understanding of the rule itself. The exceptions may even prove to be extensions of the rule, and such is the case with exceptional mutants in the complementation phenomenon.

Complementation has been observed between different mutants whose defects involve the production of nonidentical subunits composing a single enzyme. This was pointed out above for the two genes involved in the production of the A and B components of bacterial tryptophan synthetase. As you recall, pair-wise combinations of different A mutants, or pair-wise combinations of different B mutants uniformly failed to complement, while every A mutant complemented every B mutant. These exceptions, then, help to clarify the relationship of one gene to one enzyme in more precise and correct terms: one gene, one polypeptide chain.

Complementation has also been observed to occur between two mutants, neither of which complements a third mutant. The possibility that the third mutation is an aberration that involves both of the other two sites of mutation can be ruled out by recombination tests. The single gene, therefore, can be further divided into subunits called complementation units. Some mutations involve one complementation unit *a*, other mutations involve complementation unit *b*, while still other mutations affect simultaneously both the *a* and *b* complementation units. The best current evidence has led to the working hypothesis that different complementation units within a single gene involve portions of the gene dictating the structure of particular sections of the polypeptide chain. Such sections might, for example, be specific alpha-helical segments of the protein. These segments of the chain may be the regions involved in protein-protein interactions between identical molecules. In other words, those genes exhibiting intragenic complementation are believed to control the production of enzymes that are polymers composed of one or more types of identical subunits. The single polypeptide chains, the monomers, either are inactive or have very limited enzymic activity. The evidence for these conclusions stems from studies of the general type discussed below.

A simplified complementation map schematizing the complementation behavior of *Neurospora* tryptophan synthetase mutants is shown in Fig. 5.3. Mutants that complement are indicated by nonoverlapping lines, while mutants that fail to complement are indicated by overlapping lines. While initially it was thought that CRM$^-$ mutants were noncomplementing mutants, it has recently been found that many of the CRM$^-$ mutants do indeed complement with other tryptophan synthetase mutants, but the complementation is very inefficient. Wild-type *Neurospora* tryptophan synthetase, isolated as a pure protein, has a molecular weight of about 120,000 and contains about 1,200 amino acid residues. The quantities of arginine and lysine are approximately equal, the total number of these two basic amino acids being about 130. One would therefore predict that, when *Neurospora* tryptophan synthetase is digested with trypsin (which splits the polypeptide chain at the arginine and lysine residues), about 130 peptide fragments would be detected on a fingerprint profile. Surprisingly, only about 48 peptides were detected. This observation, plus several lines of physical evidence, suggests that the *Neurospora* tryptophan synthetase is composed of several similar polypeptide chains. More evidence will be needed to establish the number and chemical nature of the monomers in the polymer structure.

It is quite possible that the individual monomers are enzymatically inactive in the indoleglycerolphosphate-to-tryptophan reaction and that only the polymer is active. In this postulation, the defective pro-

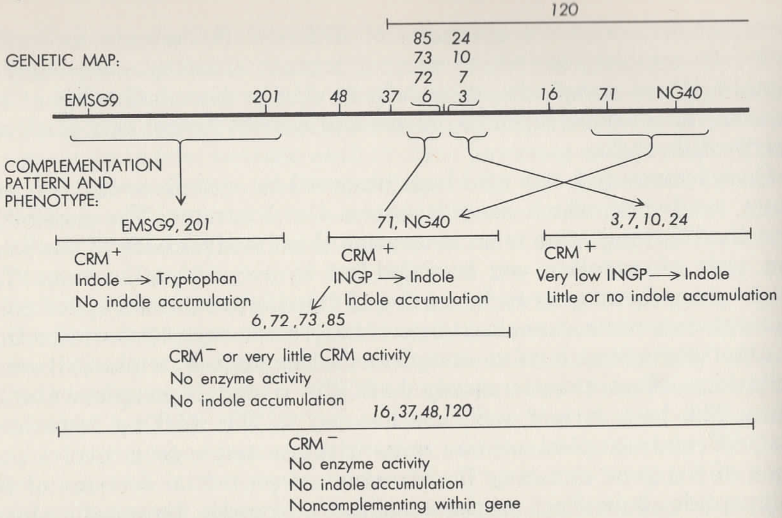

120
GENETIC MAP:
85 24
73 10
72 7
EMSG9 201 48 37 6 3 16 71 NG40
COMPLEMENTATION PATTERN AND PHENOTYPE:
EMSG9, 201
CRM +
Indole ⟶ Tryptophan
No indole accumulation
71, NG40
CRM +
INGP ⟶ Indole
Indole accumulation
3, 7, 10, 24
CRM +
Very low INGP ⟶ Indole
Little or no indole accumulation
6, 72, 73, 85
CRM⁻ or very little CRM activity
No enzyme activity
No indole accumulation
16, 37, 48, 120
CRM⁻
No enzyme activity
No indole accumulation
Noncomplementing within gene

Fig. 5.3. Genetic and complementation maps of the tryptophan synthetase gene of *Neurospora crassa*. The heavy bar represents the chromosome; the numbers above the bar are the sites of mutation, mapped by recombination tests, of particular mutants designated by letters and numbers. Included is one mutant (*120*) that is a multisite or extended mutation, encompassing a number of individual sites; it is presumed to be an aberration, for example, a deletion. A number of independent mutations cluster near one spot (*6, 3,* etc.).

The complementation pattern and the characteristics of the mutants are listed below the genetic map. The complementation map is read vertically: those mutants whose lines overlap fail to complement; those mutants whose lines do not overlap do complement. Note that there is good correspondence between behavior in complementation, properties of the protein elaborated by the mutant, and genetic localization on the chromosome. The two maps are not strictly colinear (note that the two arrows at the right side of the figure cross each other; were the two maps strictly colinear, the lines would not cross). However, one can join the two ends of the genetic map, making it a circle, and attain colinearity with the complementation map. Similar relationships are found in other genes of *Neurospora*.

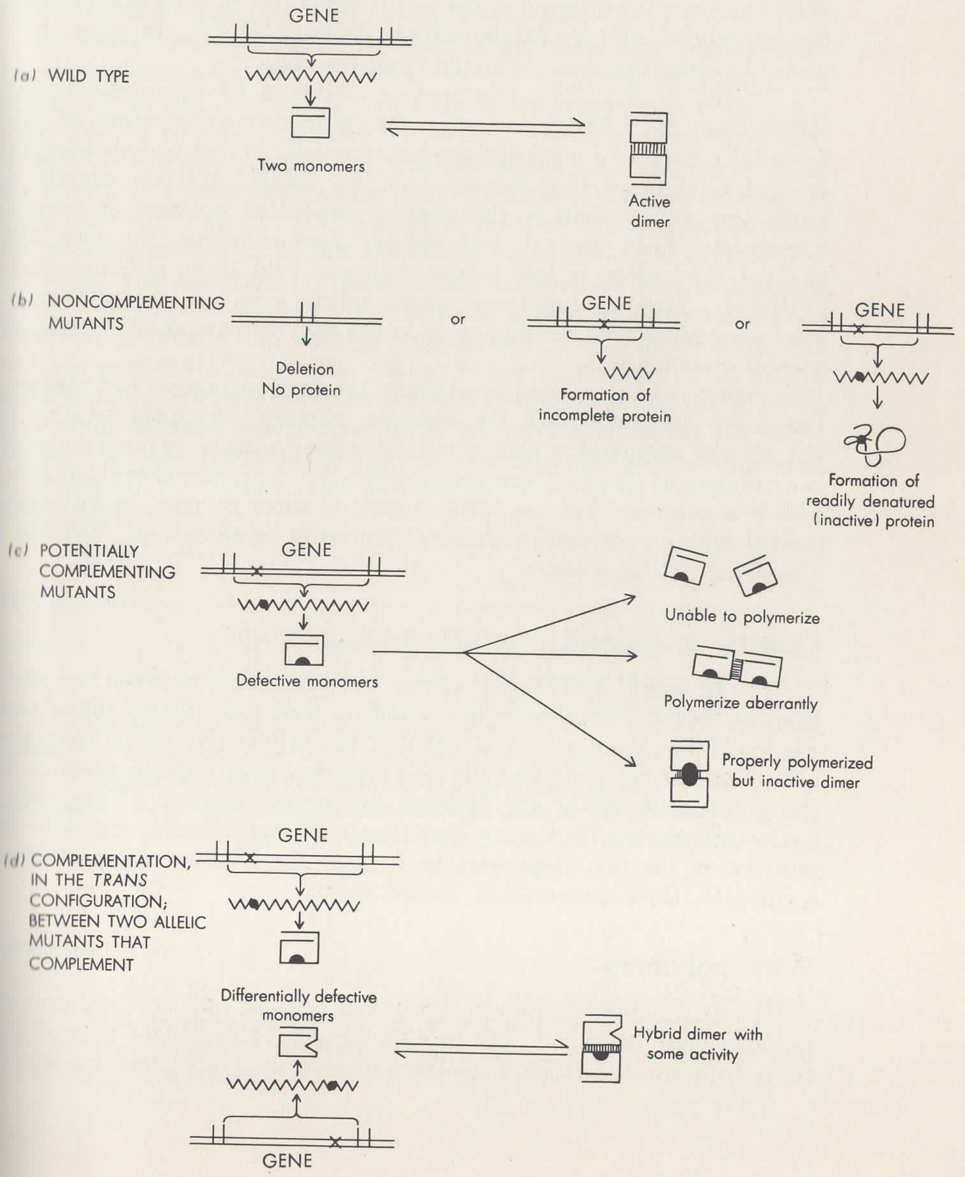
Note that noncomplementing mutants are scattered through the gene and that these are uniformly CRM⁻ mutants and include the multisite mutant, *120*. This pattern is typical of many other gene loci. Based on data of A. M. Lacy and D. M. Bonner, *Proc. Nat. Acad. Sci. U.S., 47* (1961), 72; Y. Suyama, A. M. Lacy, and D. M. Bonner, *Genetics, 49* (1964), 135; S. Kaplan, Y. Suyama, and D. M. Bonner, *Genetics, 49* (1964), 145; and A. M. Lacy, in a personal communication.

tein produced by a CRM$^+$ mutant may represent the aggregation of structurally altered subunits into a catalytically ineffective polymer (Fig. 5.4c). Perhaps, when complementation takes place between two mutants, it is the result of the interaction of differentially defective monomers and the subsequent aggregation into an active polymer (Fig. 5.4d). As mentioned earlier, there is some evidence that CRM$^-$ mutants also form a protein product that appears to be a grossly abnormal CRM, not recognizable by the usual methods of CRM assay. The CRM$^-$ phenotype may reflect damage to the polymeric structure of the protein by a substitution of the wrong amino acid at a site in the polypeptide chain that is critical for the association of chains into a structure recognizable as a CRM (Figs. 5.4b and 5.4c). The recent finding that many so-called CRM$^-$ tryptophan synthetase mutants of *Neurospora* complement, although very poorly, certainly argues that an active polymer may be formed by these mutants under certain conditions.

Complementation in vitro

Extracts of histidine B mutants all lack dehydrase activity, and some also lack phosphatase activity, the second function of the

Fig. 5.4. Protein-protein interactions of similar polypeptide chains, giving rise to active enzyme.


GENE
(a) WILD TYPE
Two monomers
Active dimer
(b) NONCOMPLEMENTING MUTANTS
Deletion
No protein
or
GENE
Formation of incomplete protein
or
GENE
Formation of readily denatured (inactive) protein
(c) POTENTIALLY COMPLEMENTING MUTANTS
GENE
Defective monomers
Unable to polymerize
Polymerize aberrantly
Properly polymerized but inactive dimer
(d) COMPLEMENTATION, IN THE TRANS CONFIGURATION; BETWEEN TWO ALLELIC MUTANTS THAT COMPLEMENT
GENE
Differentially defective monomers
Hybrid dimer with some activity
GENE

bifunctional protein (Fig. 4.5). However, by mixing extracts of mutants belonging to different complementation units (as defined by the *trans* test in vivo), dehydrase activity can be restored. Restoration of dehydrase activity requires reducing agent and divalent cation, which are expected to influence protein-protein interactions. Thus, the indications are that the interaction of polypeptide chains, pictured in Fig. 5.4d as occurring in vivo during complementation, is reproduced in vitro.

Some mutants that complement weakly in vivo fail to complement in vitro; mixed extracts fail to reconstitute enzyme activity. These mutants also lack phosphatase activity and CRM in vitro. Presumably, the complementation noted in the *trans* test in vivo is caused by the interaction of very labile, altered proteins, which are completely unstable when they are extracted from the cell.

As noted in Fig. 5.4d, the enzyme obtained by complementation differs from the wild-type enzyme. Its distinctive composition, of differentially defective polypeptide chains, results in distinctive catalytic and physical properties. For example, the hybrid polymer usually is much less active than is the enzyme consisting entirely of normal monomers. When normal and mutant monomers are together, the enzyme very often is less active than the wild type, although with particular, rare combinations, some hybrid enzyme molecules may have enzymic activity comparable to the enzyme composed entirely of normal monomers.

Certainly, at the present time, many facets of intragenic complementation are not understood. For example, intragenic complementation is not always observed within a cistron whose normal enzyme product is a polymer of identical monomers. Wild-type pyrophosphorylase (Fig. 4.5) is a polymer. Yet, mutants in gene *G* show no complementation, even though an adequate number of pair-wise combinations (450 combinations) has been tested.

Genetic maps and complementation maps

In many cases, there is some correspondence between the positions of mutants on genetic maps and on complementation maps (see Fig. 5.3). The correlations of position are believed to result from the colinearity of the gene and the polypeptide chain. Certain sections of the polypeptide chain are dictated by certain sections of the polynucleotide chain. That some mutations do not strictly correlate in position on the two maps may be because the final, active protein is a complex, three-dimensional, folded chain.

Why polymers?

In Chapter 2, the interactions of tobacco mosaic virus protein subunits were discussed. The protein monomers aggregate around the RNA in a specific fashion, to form a comparatively giant structure, the intact virus rod (Fig. 2.12). It is estimated that only a relatively small amount (8 per cent) of the total genetic information of tobacco mosaic virus is necessary to code for the formation of the monomer. Nevertheless, this amount is sufficient to permit about 2,200 monomeric units to polymerize into the coat that protects all the labile RNA. The advantage of the subunit building-blocks system is clear; the polymerization process allows the maximum of protection (survival value) for the minimum amount of hereditary baggage. Or, this could be expressed in other words as the conservation of genetic information.

The polymer-type structure of protein molecules also permits, as we have seen, the repair of many genetic defects by complementation mechanisms involving protein-protein interaction.

The selective advantages that enzyme polymers may bestow upon higher organisms can be illustrated by the following possibilities. A heterozygous cell, containing two slightly different genetic elements, may contain several polymeric forms of one enzyme. Possibly, each form of the enzyme is better suited to one particular environment. Also, the activities of polymeric and monomeric units of the enzyme may be quite different. Since the polymerization reaction can be brought about by environmental factors, this allows some environmental control over the rates and direction of intracellular metabolism. For instance, a number of enzymic activities are known to appear suddenly during the birth of a mammal. The enzyme proteins are present before birth, but are nonfunctional. The drastic change in the environment that occurs at birth could be conceived as causing the alteration of polymeric forms of certain enzymes and consequently of certain activities. In other words, the monomer-polymer transition systems may provide one type of biological regulation: some of these aspects will be discussed in a later chapter.

References

Ames, B. N., and P. E. Hartman, in *The Molecular Basis of Neoplasia.* Austin: University of Texas Press, 1962. Review of genetic and biochemical studies on the histidine biosynthetic enzymes in *Salmonella,* including analysis by complementation.

Carlson, E. A., *Quart. Rev. Biol., 34* (1959), 33. Review of gene loci in *Drosophila,* including discussion of complementation results.

Fincham, J. R. S., and A. Coddington, *J. Mol. Biol., 6* (1963), 361. A report that presents experimental evidence bearing on interpretation of the phenomenon of intragenic complementation.

Questions

5.1. Based upon Fig. 5.2, set up a *cis* genetic test with a diploid organism. Mention the organism and tell why it was selected for the test.

5.2. Set up a *cis* genetic test for a normally haploid organism. Explain your choice of organism.

5.3. Consult the library and write a short report on the properties of hemoglobin that lead to the view that the functional protein is a tetramer.

5.4. The predominant hemoglobin in adult man is hemoglobin A, a tetramer of two alpha and two beta chains. About 2.5 per cent of normal adult hemoglobin (hemoglobin A2) differs from hemoglobin A in that it is composed of two alpha and two delta chains. The predominant hemoglobin of the human fetus is hemoglobin F, composed of two alpha and two gamma chains. Normal children gradually shift to almost exclusive production of hemoglobins A and A2 within a few months after birth. Individual genes are responsible for each type of polypeptide chain.
(*a*) Hemoglobin A2 has a higher affinity for oxygen than do the other hemoglobins. State how this could be a case of gene evolution.
(*b*) Individuals have been found who produce tetramers composed of four beta chains or, in other cases, four alpha chains. What reasons could you give to explain the failure to find such tetramers in normal individuals? Would you expect the abnormal tetramers to be active? Explain in chemical and in biological terms why you think so. Besides the tetramers comprising four identical chains, what other hemoglobins would you expect to find in these individuals? Why?
(*c*) Certain regions of the hemoglobin polypeptide chains interact to give rise to the functional tetramers. Explain the distribution of mutations you would expect to find: (1) affecting the portion of the molecule, (2) affecting polymerization, (3) located in external portions of the polypeptide chain not directly involved with polymerization or with catalytic activity.

5.5. Read the sections of McKusick's *Human Genetics* in this series that deal with the structure of normal and abnormal hemoglobins. Reorganize the material into a one-page synopsis of the critical features of the genetics and protein chemistry of the hemoglobins.

5.6. In the literature find a published complementation map for some gene other than those mentioned in this chapter and interpret its significance relative to the structure of the enzyme concerned.

5.7. Heterozygotes are often found to have a selective advantage over homozygotes. Elaborate on what role you think complementation *between* genes and complementation *within* genes might play in this phenomenon. Give the reasoning behind your arguments.

Six

Secondary Consequences of Gene Mutations

The primary effect of a gene mutation is to cause an alteration in the amino acid sequence of a protein molecule. The protein is synthesized in a system that includes mRNA as the specific template. Therefore, we can say: one mutant gene, one mutant mRNA, one mutant polypeptide chain. This seemingly simple primary event causes a multitude of secondary effects that ultimately result in the mutant phenotype. In the vast majority of mutant phenotypes that have been studied, only the secondary effects have been analyzed. An understanding of the secondary effects of gene mutations is instructive, for they have provided us with much information about normal physiological processes. Because individual biochemical reactions usually are components of more complex anabolic or catabolic sequences in cellular metabolism, relatively few single biochemical reactions in the cell can be disrupted without affecting other metabolic steps.

Localization of genetic blocks

The most immediate effect of the alteration of an enzyme is a change in the efficiency of a particular step in a metabolic sequence. If the mutant enzyme is inactive, we say that the reaction sequence is blocked, or that a genetic block is present. We have seen this effect in mutants unable to synthesize the amino acids tryptophan and histidine (Figs. 4.1 and 4.5).

If the genetic block in a microorganism shuts off the synthesis of an essential metabolite, thereby creating a nutritional requirement, the metabolite often can be supplied to the organism from the environment. For example, histidineless mutants of microorganisms fail to grow on minimal medium, but they can grow with a normal growth rate on a medium supplemented with L-histidine.

The growth response of nutritional mutants to known compounds can sometimes provide information about the position of genetic blocks in a reaction sequence and, to a lesser extent, provide clues to the identity of possible biochemical intermediates. For example, histidine-requiring mutants of bacteria are of two types: those that grow on either L-histidine or L-histidinol, and those that respond only to L-histidine (Fig. 6.1). The latter are blocked in reaction 10, the conversion of histidinol to histidine. Mutants that have genetic blocks earlier in the pathway (steps 1 through 9) can grow on histidinol, because they still retain reaction 10. In contrast, no *Neurospora* histidineless mutants can grow on L-histidinol, even though this is an intermediate in histidine biosynthesis; L-histidinol is not normally permeable in *Neurospora,* although the organism can gain the ability to utilize L-histidinol through mutation. Neither *Neurospora* nor *Salmonella* can take up from the medium the phosphorylated intermediates in the pathway: hence, they fail to grow on these compounds. Furthermore, these organisms cannot phosphorylate imidazoleglycerol or imidazoleacetol to the active phosphorylated derivatives (Fig. 6.1).

Thus, growth on known compounds is sometimes a helpful tool, but its use in biochemical and genetical analyses requires cautious interpretation. Another case in point involves tryptophan biosynthesis (Fig. 6.2). Growth responses to anthranilic acid and indole can differentiate mutants blocked in the various steps of tryptophan biosynthesis. However, the fact that some tryptophan synthetase mutants respond to indole does not necessarily indicate that free indole is an actual intermediate in the biosynthetic pathway. In fact, for a time, the accumulation of indole by bacterial tryptophan synthetase B protein mutants and the growth of bacterial tryptophan synthetase A protein mutants on indole obscured the true nature of the physiological reaction and the relationship between the A and B protein subunits that compose tryptophan synthetase (for review, see Chapters 4 and 5). Similarly, growth of bacterial histidineless mutants on D-histidine, imidazole lactic acid, and imidazole propionic acid does not mean that these compounds are involved in the pathway of L-histidine biosynthesis in normal cells (Fig. 6.1).

Fig. 6.1. Pathway of L-histidine biosynthesis in *Salmonella* and some accessory reactions. The physiologically important pathway of biosynthesis is indicated by solid arrows. Accessory reactions, which are presumed from accumulation and growth data, are indicated by the broken arrows. Reaction steps are designated by numbers. The circled letter under each enzyme designation is the gene locus involved in dictating the structure of that enzyme (consult Fig. 4.5 for a detailed genetic map).

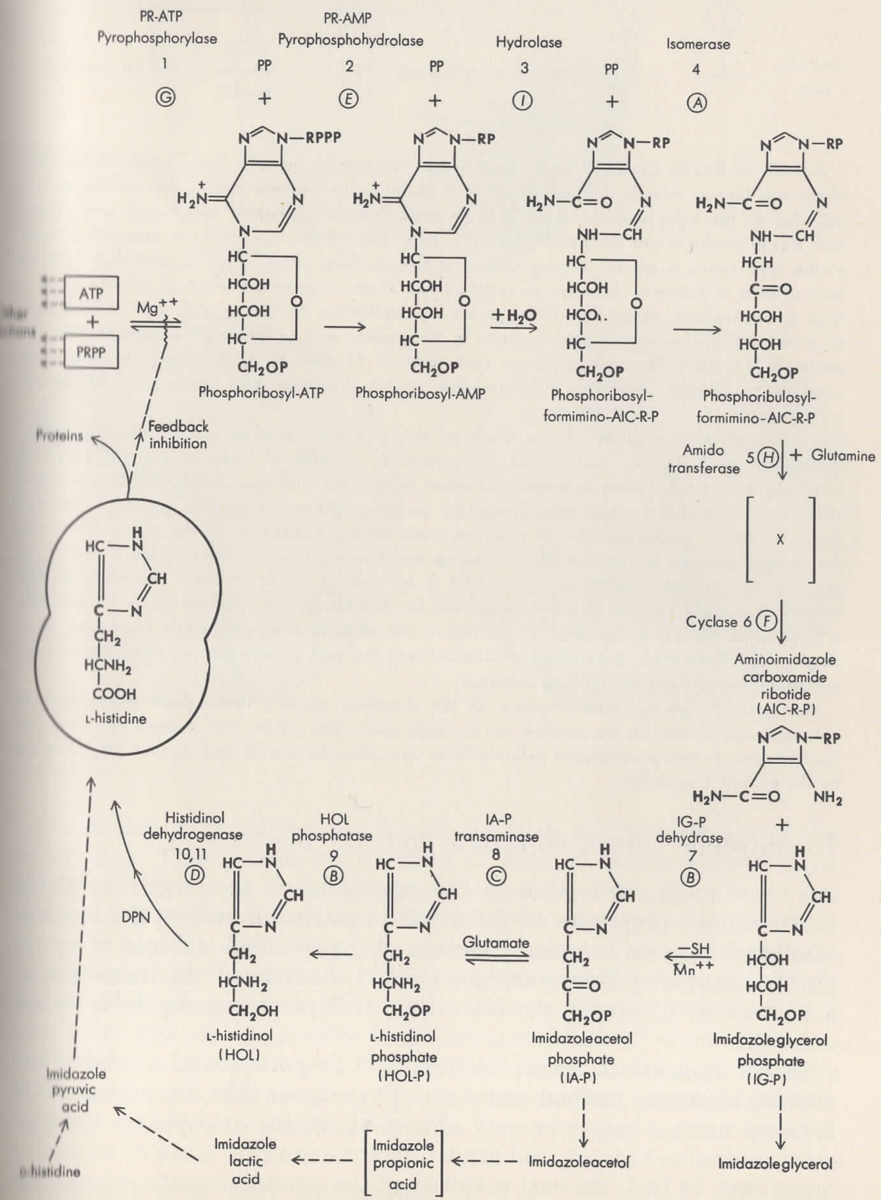
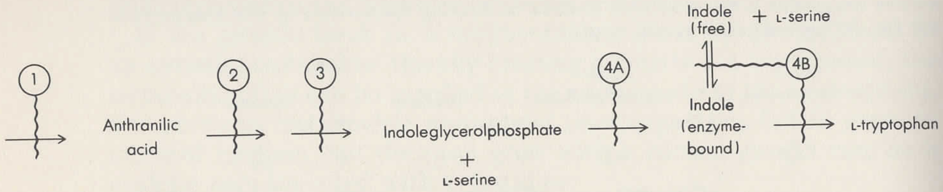

PR-ATP Pyrophosphorylase
1
Ⓖ
PP
+
PR-AMP Pyrophosphohydrolase
2
Ⓔ
PP
+
Hydrolase
3
Ⓘ
PP
+
Isomerase
4
Ⓐ
ATP
+
PRPP
Mg++
+H₂O
Phosphoribosyl-ATP
Phosphoribosyl-AMP
Phosphoribosyl-formimino-AIC-R-P
Phosphoribulosyl-formimino-AIC-R-P
Feedback inhibition
Proteins
L-histidine
Amido transferase
5 Ⓗ
+ Glutamine
X
Cyclase 6 Ⓕ
Aminoimidazole carboxamide ribotide (AIC-R-P)
Histidinol dehydrogenase
10,11
Ⓓ
DPN
HOL phosphatase
9
Ⓑ
IA-P transaminase
8
Ⓒ
Glutamate
IG-P dehydrase
7
Ⓑ
—SH
Mn++
L-histidinol (HOL)
L-histidinol phosphate (HOL-P)
Imidazoleacetol phosphate (IA-P)
Imidazoleglycerol phosphate (IG-P)
Imidazole pyruvic acid
D-histidine
Imidazole lactic acid
Imidazole propionic acid
Imidazoleacetol
Imidazoleglycerol

Fig. 6.2. Growth properties and cross-feeding by tryptophan-requiring mutants of bacteria. A more detailed presentation of the biochemical pathway is shown in Fig. 4.1.

Biochemical Pathway.



Growth on Known Compounds. A large number of mutant bacteria (i.e., 1,000–10,000,000) either are spread directly on the surface of a glucose-salts medium lacking tryptophan or are included in the agar medium when it is poured. Crystals of known compounds are placed with sterile toothpicks at the points indicated. After incubation, the ability to respond to each of the compounds is scored. Photograph *a* on the opposite page shows a petri dish covered with a lawn of bacteria defective in reaction 4A. There is some inhibition of growth at very high concentrations of indole, at the point of application of the crystal. This mutant fails to grow on anthranilic acid but responds to tryptophan and to indole. You can imagine the patterns of growth response obtained with mutants blocked in other reactions. This is an excellent qualitative procedure for screening growth responses and interactions of various compounds.

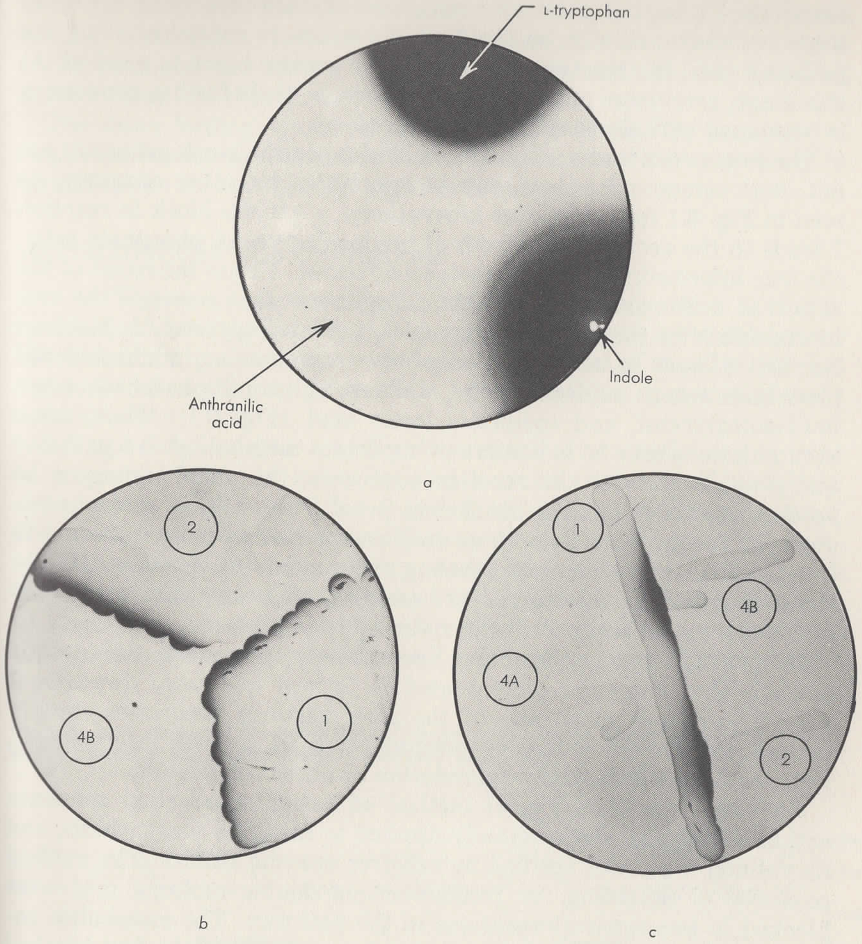
Cross-Feeding. The ability of one bacterial strain to feed another can be readily demonstrated by the following methods. In photograph *b*, cultures of three different tryptophan-requiring strains have been separately streaked on glucose-salts agar lacking tryptophan. The streaking, in a side-by-side boomerang-like pattern, allows compounds excreted by one strain to diffuse to the position of another strain during incubation of the plate. Note that the mutant blocked in reaction 4B is feeding the mutants blocked in reactions 1 and 2. In contrast, the mutant blocked in reaction 4B is not fed by either of these two mutants. The mutant blocked in reaction 2, while being fed by that blocked in reaction 4B, is nevertheless able to feed the mutant blocked in reaction 1. All of these interactions can be scored on a single plate. Sometimes, suboptimal enrichment with the end product (in this case, tryptophan) enhances the clarity of the feeding response.

Photograph *c* shows a modification of the streaking pattern that allows interactions of four different strains to be studied on a single petri dish. Note that mutant 4A does not feed mutant 1; the accumulated intermediates are phosphorylated and cannot be taken up by the mutant 1 bacteria.

Incomplete versus complete genetic blocks

The characteristics of the genetic block are largely referable to the unique properties of the altered protein formed by the mutant. The block may be complete, because of the complete absence of active protein, or it may be incomplete (leaky), because of the formation of a protein with altered structure that still possesses slight biological activity.

Leaky mutants are quite common, and they are found whenever the mutant screening method can detect phenotypes that are intermediate between normal and extremely abnormal, as, for example, in eye-pigment mutants of insects and growth factor–requiring mutants of microorganisms. In fact, the vast majority of the mutants observed in higher organisms are of the leaky type; a minority of mutants exhibit extreme phenotypes, sometimes resulting from deletion of the gene (see *Cytogenetics* and *The Mechanics of Inheritance* in this series).



a

b

c

Occasionally, genotypes are found in which all mutants are of the leaky type. This is true for mutants of gene *H*, which appear to be blocked in step 5 of histidine biosynthesis (Fig. 6.1). The reason lies in the nature of the reaction. In the presence of enzyme, the reaction proceeds rapidly with very low concentrations of normal substrate (glutamine); in the absence of enzyme, the reaction proceeds slowly and spontaneously when high concentrations of ammonium ion are available. Because relatively high concentrations of ammonium ion are present in all the bacteria, the reaction occurs at a suboptimal rate in the mutants in the absence of the specific enzyme, thereby facilitating slow synthesis of histidine.

Accumulations and precursors

A single genetic block may be pictured as a specific damming of the flow of metabolites at one spot in a sequence of biochemical reactions. Up to that particular step, the enzymes catalyzing the other steps function normally, so the flow of precursors continues along the pathway until the reaction affected by the genetic block is reached. At this stage, precursors may accumulate because their further conversion is prevented entirely or markedly reduced.

The precursors that accumulate behind a genetic block generally are not inert compounds; they may be substrates for other enzymes. As seen in Fig. 6.1 for L-histidine biosynthesis, a genetic block in reaction 7 leads to the accumulation both of imidazoleglycerol phosphate ester, the true intermediate, and of imidazoleglycerol. This is the result of the action of a ubiquitous nonspecific phosphatase that converts the true intermediate to the secondary product, imidazoleglycerol. A block at reaction 8 leads to accumulation of five compounds: imidazoleacetol phosphate ester, imidazoleacetol, imidazoleglycerol phosphate ester, imidazoleglycerol, and imidazole lactic acid (Fig. 6.1). Thus, while accumulations may be indicative of important metabolites in a pathway and provide clues to the reaction sequences, the mere finding of an accumulated compound is insufficient proof of its role as a true intermediate. Other criteria—such as enzyme conversion to end product in cell-free extracts, compound labeling experiments using radioactive or stable isotopes, and so forth—are essential to assure that the reconstructed reaction sequence is physiologically important.

Frequently, the intermediate immediately preceding the genetic block is the compound accumulated in highest quantity. However, a word of caution: often one of the side reactions progresses rapidly enough to prevent extensive accumulation of a true intermediate, while causing substantial accumulation of other compounds.

An elegant microbiological method of assigning reaction sequence within a pathway was originally applied to histidine biosynthesis, and its validity was later verified by cell-free enzyme studies. The method consisted of obtaining, by recombination, double mutants that were blocked in two separate reactions in the pathway. The metabolite accumulated by the double mutant was characteristic of the first blocked reaction. The metabolite normally accumulated as a result of the second genetic block was not found, because the supply of its precursor had been shut off at the earlier reaction step.

Accumulations and feeding

Mutants with genetic blocks that preclude growth on minimal medium will grow only on intermediates that are located beyond the genetic block, or on compounds that can be converted to intermediates. This observation leads one to expect that mutants blocked in later steps in a biochemical reaction sequence would accumulate intermediates that satisfy the growth requirement of mutants blocked in earlier reactions. For example, mutants blocked in reaction 10 of histidine biosynthesis (Fig. 6.1) should accumulate L-histidinol, which can be used by organisms blocked in earlier reactions. Indeed, such cross-feeding phenomena (syntrophism) are found with many mutants. Other examples are shown in Fig. 6.2, where methods for detecting cross-feeding are described.

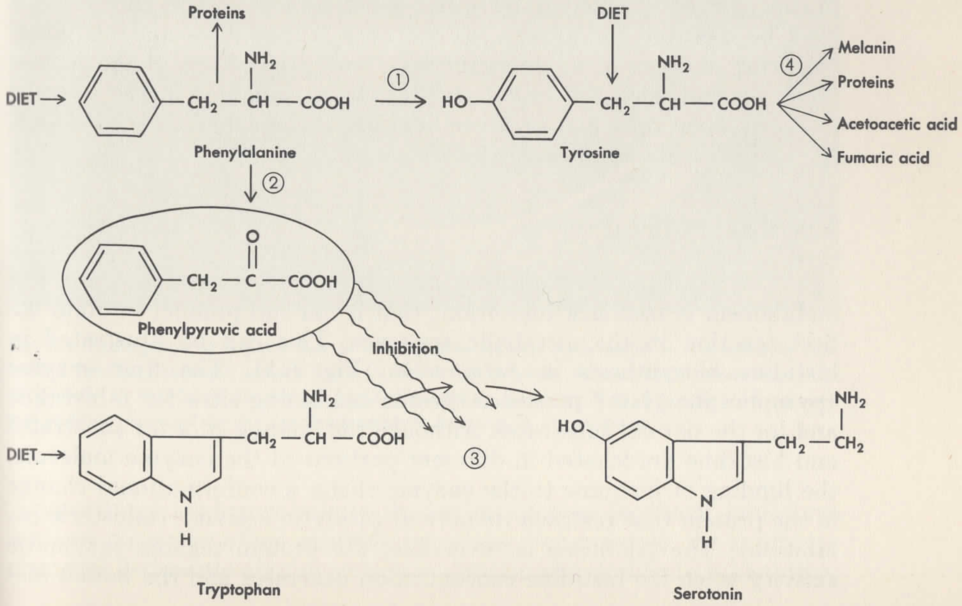
The cross-feeding technique is useful in ordering the sequence of suspected biochemical reactions as well as in detecting and aiding in identification of biologically significant compounds.

Accumulations and inhibitions

The accumulation by a mutant of an otherwise normally metabolized intermediate might be expected to have some deleterious effect on the organism. Many examples are known of metabolites that are toxic when they are present at concentrations in excess of their normal levels.

Let us briefly consider the effect in humans of accumulations in front of a genetic block. A more detailed account of inherited metabolic disorders in humans will be found in Victor A. McKusick's book, *Human Genetics,* in this series. Humans cannot synthesize phenylalanine; they must rely on a dietary intake of this essential amino acid. The excess phenylalanine consumed in the diet is normally converted to tyrosine (reaction 1, in Fig. 6.3), which is metabolized through a series of reactions to acetoacetic acid and fumaric acid. However, some

Fig. 6.3. Some aspects of the metabolism of phenylalanine in the human being. The numbers signify particular biochemical reactions discussed in the text.



human mutants possess genetic blocks in reaction 1. Such persons retain high blood levels of phenylalanine and excrete phenylpyruvic acid and related compounds into the urine. Intracellular phenylpyruvic acid accumulates in sufficient quantity to inhibit the conversion of tryptophan to serotonin (reaction 3, in Fig. 6.3). Current evidence indicates that serotonin is essential for normal mental processes. Phenylketonuric persons are usually mentally retarded, presumably because of the secondary effects of the genetic block.

Here, then, is an example of a chain of events arising from a single genetic defect. This defect is expressed initially as a metabolic lesion, and eventually it causes a gross phenotypic alteration, mental retardation. Because tyrosine is present in the diet, the intermediates beyond the genetic block are provided. The deleterious effect of the genetic block, then, is caused by the metabolites that accumulate.

End-product deficit

Metabolites beyond the genetic block may not be essential to normal growth and development, or they may be supplied by other reactions. For example, persons blocked in reaction 1 (Fig. 6.3) have less tyrosine available than do normal persons and thus are more lightly pigmented (contain less melanin) than are normal individuals. Presumably, the human phenylketonuric, blocked in reaction 1, would show more severe signs of tyrosine deficiency were his diet deficient in this amino acid.

In contrast, the absence of the end product is the most important phenotypic effect elicited by other genetic blocks. The end product may be essential for growth, as with the tryptophan- and histidine-requiring mutants of microorganisms. Hence, the effects of the genetic block depend upon the essentiality of the metabolites in the affected pathway, their roles in over-all cellular metabolism, and their availability and utilizability from the environment.

Feedback inhibition

One of the most striking examples of the regulation of cellular metabolism is feedback inhibition; that is, an end product inhibits the first reaction in the metabolic sequence. This can be illustrated in histidine biosynthesis in *Salmonella* (Fig. 6.1). The first enzyme (pyrophosphorylase) possesses specific combining sites for L-histidine and for the normal substrates. Although the binding sites for substrates and histidine are located in different portions of the enzyme molecule, the binding of histidine to the enzyme elicits a conformational change in the protein that restricts its normal catalytic activity (allosteric inhibition). The inhibition is reversible; the protein regains enzymatic activity when the histidine concentration decreases and the bound histidine gradually dissociates from the enzyme. Feedback inhibition thus acts as a shut-off valve; it allows the cell to tap precursors and energy from the metabolic pool at a rate essential for efficient biosynthesis—and no faster. The concentration of free histidine normally present in *Salmonella* is enough to inhibit partially pyrophosphorylase activity. That is, on the average, more total enzyme is present than is functioning to manufacture histidine.

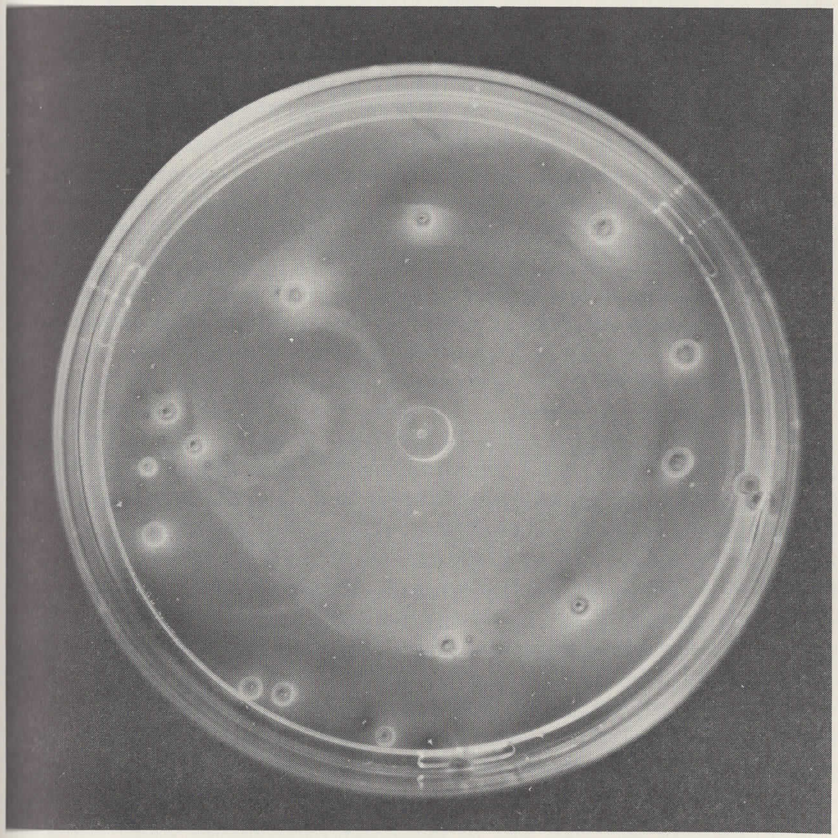


Fig. 6.4. Photograph of a petri dish of mineral-salts medium upon which a few bacteria containing feedback-resistant enzyme have formed large colonies. Around each colony is a halo of growth of the large surplus of histidine-requiring bacteria. They are using for growth the histidine excreted by the feedback-resistant bacteria in the colonies.

Sometimes analogues mimic the action of true feedback inhibitors without replacing other valuable functions. For example, the synthetic chemical 2-thiazole alanine, which is structurally similar to L-histidine, is a feedback inhibitor, but it cannot be activated or incorporated into protein. It inhibits bacterial growth because it shuts off production of histidine but does not replace histidine for protein synthesis. Mutants resistant to 2-thiazole alanine also resist feedback inhibition by the normal intracellular concentrations of histidine. As a result, in the absence of thiazole alanine, they make excess histidine, which is excreted into the medium (Fig. 6.4). These feedback-resistant mutants map in the pyrophosphorylase gene (gene *G* in Fig. 4.5). They form an altered pyrophosphorylase enzyme that retains its catalytic activity but is not inhibited by low concentrations of histidine or 2-thiazole alanine.

Feedback inhibition can be demonstrated readily in vitro by inhibition of enzyme activity. It also can be demonstrated by two methods in vivo. After bacteria grow in medium containing radioactive (C^{14}) glucose as sole carbon source, all of the carbon-containing compounds of the cell become labeled with C^{14} at identical specific activities. When C^{12} (nonradioactive) L-histidine also is present during growth, all carbon compounds still contain high specific activities of C^{14} except the histidine that has been incorporated into protein. The exogenous histidine is thus preferentially utilized, although cell extracts still contain the enzymes involved in histidine biosynthesis. High preferential utilization of this type is suggestive of feedback inhibition, but must be further confirmed by in vitro enzyme studies with cell-free extracts.

The second in vivo method for detecting feedback inhibition is to observe the time course of accumulation of intermediates by nutritional mutants (Fig. 6.5). In histidine-requiring mutants growing in medium containing a small amount of L-histidine, the potentially competent reactions in the biosynthetic pathway of the cell are arrested until the histidine in the medium is utilized completely. When the histidine supply is depleted, feedback inhibition ceases and the bacteria stop growing. The previously inhibited reactions in the histidine biosynthetic sequence now proceed, and the intermediates characteristic of the genetic block begin to accumulate.

The process of inhibition can be dissociated from growth and protein synthesis by a refinement of the method shown in Fig. 6.5. This requires suspension of mutant bacteria in an environment that prevents protein synthesis. The bacteria accumulate precursors until the end product is added, at which time further accumulation ceases. In this manner, the ability to accumulate is demonstrated before the feedback inhibitor is added.

The phenomenon of feedback inhibition demonstrates that the concentration of the end product of a metabolic pathway can control the rates of biochemical reactions within the pathway. Therefore, in a mutant deprived of the end product, the reaction that is normally inhibited can be released from inhibition and attain a rapid, uncontrolled rate. The widespread occurrence of feedback inhibition in microorganisms as well as in mammals suggests that sensitivities to feedback inhibitions may play important roles in the secondary consequences of many genetic blocks.

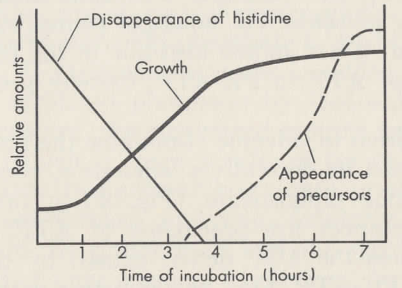


Fig. 6.5. Idealized curves showing the effects of feedback inhibition on accumulations of intermediates in a histidine-requiring mutant of *Salmonella*. The histidine-requiring bacteria take up histidine from the medium ("Disappearance of histidine") and, after a short lag period, grow exponentially until essentially all of the histidine is exhausted from the external environment, at which time growth ceases ("Growth"). Because of feedback inhibition, the intermediates before the genetic block do not accumulate until the histidine concentration has reached an extremely low level.

Competition for substrate

The severe consequences of release from feedback inhibition are exemplified in the behavior of histidine-requiring bacteria. Included in Fig. 6.1 is a diagram of the feedback-sensitive reaction in histidine biosynthesis: the conversion of adenosine triphosphate (ATP) and phosphoribosylpyrophosphate (PRPP) to phosphoribosyladenosine triphosphate (PR-ATP). As indicated in the diagram, the two substrates, ATP and PRPP, participate in a number of other reactions; for example, ATP supplies the energy for amino acid activation (Chapter 3). Therefore, the release of feedback inhibition and excessive accumulation of worthless biochemical intermediates behind a genetic block might result in an excessive drain of the energy-rich substrates necessary for other biochemical reactions in the cell. For example, many histidine-requiring bacteria have absolute requirements for L-histidine. However, some leaky mutants grow very slowly on minimal medium lacking histidine. They synthesize histidine at a slow rate and use it preferentially for protein synthesis. Thus, the endogenous histidine is insufficient to shut down the pathway by feedback inhibition. The result is an abnormally high activity of the pyrophosphorylase enzyme that catalyzes the utilization of ATP and PRPP. This ATP is converted rapidly to PR-ATP and thence to other intermediates, which pile up before the leaky genetic block and are utilized only very slowly. This leads to a severe depletion of the free ATP supply, vital for other energy-requiring reactions of the cell. The subsequent addition of excess L-histidine to the medium restores normal growth, because it limits the conversion of ATP to PR-ATP and it is also used in protein biosynthesis. The histidine analogue 2-thiazole alanine also supports growth of the leaky mutants. The analogue cannot substitute for histidine in proteins, but it can mimic histidine in the feedback inhibition of the conversion of ATP to PR-ATP, thereby preventing ATP deficiency.

Finally, the addition of adenine stimulates the growth of the leaky mutants. Adenine and its derivatives, adenosine, adenosine monophosphate, and adenosine diphosphate (Fig. 3.1a), probably have two effects. These substances are precursors of ATP. The addition of adenine thus restores the ATP deficit caused by the excessive conversion of ATP to PR-ATP. Also, by competing with ATP for attachment to pyrophosphorylase, the adenine derivatives may reduce the enzyme activity and prevent further loss of free ATP.

As you must have already concluded, the genetic block can seriously disturb the fine balance among several biochemical pathways, and because many biochemical pathways interconnect, competition for substrate can be an important secondary consequence of a genetic block.

Duplicate enzymes and channeled pathways

Frequently, one finds a key intermediate, one that is common to several different pathways. We have just discussed ATP and PRPP, which are used in a variety of reactions.

In *Neurospora,* one enzyme catalyzes the synthesis of carbamyl phosphate to be used for arginine biosynthesis and a presumed second enzyme catalyzes the same biochemical reaction for pyrimidine biosynthesis (Fig. 6.6). The first enzyme is sensitive to feedback inhibition by arginine, and it appears likely that the presumed second enzyme is sensitive to feedback inhibition by pyrimidines.

If the carbamyl phosphates from the two pathways were interchangeable, mutants blocked in carbamyl phosphate synthesis might not be found normally; the one remaining enzyme could supply enough carbamyl phosphate for both reaction sequences. A cell that was mutant for the enzyme in one of the pathways might be revealed only in the presence of the feedback inhibitor appropriate for the other pathway. The two carbamyl phosphate pools ordinarily are not interchangeable, however.

Neurospora mutants that are blocked only in the synthesis of the pyrimidine-specific carbamyl phosphate are pyrimidine-requiring mutants. However, their requirement for pyrimidine is greatly alleviated by a second, independent mutation. Because the original mutant phenotype is suppressed, this second mutation is termed a suppressor mutation. It decreases the rate of conversion of carbamyl phosphate and ornithine to citrulline (Fig. 6.6). The suppressed mutant still can synthesize sufficient arginine for normal growth, in spite of possessing only 2 to 3 per cent of the normal enzyme activity. In the process, the mutant accumulates excess carbamyl phosphate before the genetic block, and this excess carbamyl phosphate is used to spare the pyrimidine requirement. If exogenous arginine is added, thereby curtailing the synthesis of carbamyl phosphate by feedback inhibition of the arginine-specific enzyme, the suppressed mutant again requires exogenous pyrimidine (which is characteristic of the original primary mutant). In this manner, two ordinarily separate, or channeled, pathways have become linked through mutation. Similarly, arginine-requiring mutants blocked in the synthesis of arginine-specific carbamyl phosphate are relieved of their arginine requirement by a second mutation that imposes an ATC deficiency (Fig. 6.6).

The presence of channeled pathways and duplicate enzymes can be seen to create special problems in genetical analysis, resolvable only through the closest biochemical scrutiny.

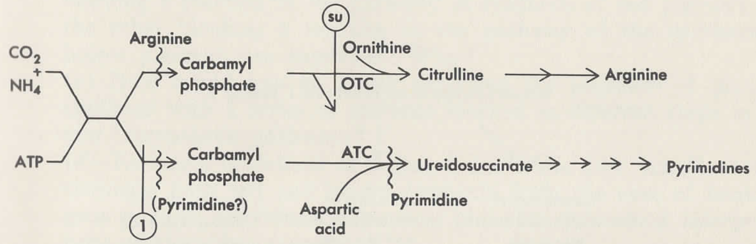


Fig. 6.6. Interrelations between pathways of arginine and pyrimidine biosynthesis. Arginine and pyrimidines have a common precursor, carbamyl phosphate. Carbamyl phosphate plus ornithine is converted to citrulline by the enzyme ornithine transcarbamylase (OTC). Carbamyl phosphate and aspartic acid are converted to ureidosuccinate, a pyrimidine precursor, by the enzyme aspartic transcarbamylase (ATC).

Consequences further removed

We have pointed out a number of metabolic abnormalities that are affected only indirectly by gene mutation. It is easy to visualize that metabolic sequences that are secondarily disturbed by mutation may, in turn, adversely affect other systems in the integrated scheme of normal metabolism.

In higher organisms, the intimate coordination of processes at the cellular, tissue, organ, and organism levels is dependent upon biochemical processes at the intracellular level. Careful analysis is required to uncover the primary source of the multiple effects manifested by various mutant phenotypes. Few mutant phenotypes in higher organisms have been traced back to the primary genetic block. The examples in this chapter may serve as models for further probing of the physiology of the normal organism and its alteration by gene mutation.

One system that has proved very fruitful in the study of higher organisms is the synthesis of the oxygen-transporting blood protein, hemoglobin. Some characteristics of the structure and inheritance of the hemoglobins are summarized in McKusick's *Human Genetics.* Figure 6.7 shows some of the main abnormalities traceable to a single cause—the replacement of glutamic acid for valine at position 6 in the beta polypeptide chain of hemoglobin. The figure shows clearly the many levels of analysis that are required to track down gene mutations in complex organisms. It also points out how a change in a nucleotide pair—effecting a single amino acid replacement—may have multiple (pleiotropic) effects.

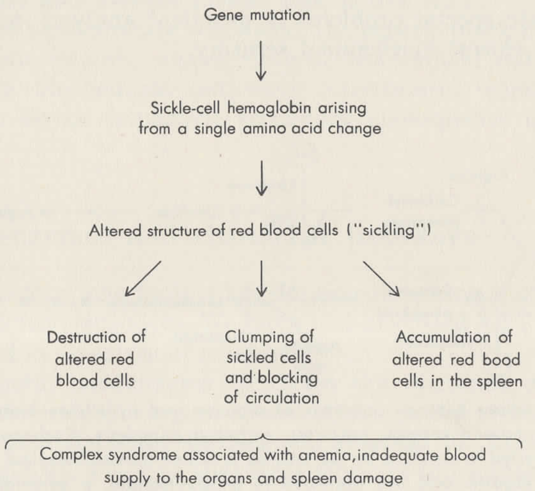

Gene mutation
Sickle-cell hemoglobin arising from a single amino acid change
Altered structure of red blood cells ("sickling")
Destruction of altered red blood cells
Clumping of sickled cells and blocking of circulation
Accumulation of altered red bood cells in the spleen
Complex syndrome associated with anemia, inadequate blood supply to the organs and spleen damage

Fig. 6.7. The widespread effects of a gene mutation in humans that alters the primary structure of hemoglobin.

References

Davis, R. H., "*Neurospora* Mutant Lacking an Arginine-specific Carbamyl Phosphokinase," *Science, 142* (1963), 1652. One in a series of original papers that have been influential in orienting our concepts of duplicate enzymes, channeled pathways, and physiological interactions.

Fruton, J. S., and S. Simmonds, *General Biochemistry,* 2nd ed. New York: John Wiley & Sons, Inc., 1958. Excellent, thorough introduction to intermediary metabolism of vintage 1958.

Umbarger, H. E., "Intracellular Regulatory Mechanisms," *Science, 145* (1964), 674. An up-to-date review with references to further reading.

Wagner, R. P., and H. K. Mitchell, *Genetics and Metabolism,* 2nd ed. New York: John Wiley & Sons, Inc., 1964. Advanced genetics text that introduces a large number of examples of gene-enzyme relationships and secondary consequences of gene mutations.

Questions

6.1. The normal eye color in *Drosophila,* red-brown, is due to admixture of two main pigments, red and brown, produced by two relatively independent biochemical pathways.

(*a*) Describe the phenotypes of two independently isolated single mutants, one blocked in the pathway leading to red pigment and one blocked in the pathway leading to brown pigment.

(*b*) What phenotypes would you expect: (1) When two mutant genes involving two separate reactions in the pathway of brown pigment synthesis are combined *cis,* and when the same two mutations are combined *trans?* Why? (2) When two mutant genes, one blocking a reaction in the pathway of synthesis of red pigment and the other blocking a reaction in the pathway of the synthesis of brown pigment, are combined? Why?

(*c*) How would you go about analyzing the sequence of pigment synthesis with a series of mutants blocked in different steps in the two biosynthetic pathways?

(*d*) Recessive mutations at a gene locus, white (*w*), simultaneously eliminate both red and brown pigments from the eyes of homozygous mutant individuals. Give two plausible theoretical interpretations of this phenomenon.

6.2. Explain pitfalls in the method of localizing genetic blocks solely on the basis of accumulations of metabolites.

6.3. From a biochemistry textbook, locate a series of biochemical reactions leading to an end product. Describe effects caused by a genetic block in the first step of the pathway. Repeat the analysis for a mutant blocked in the last step of the pathway.

6.4. Why do organisms possess some duplicate enzymes with very similar or identical catalytic activities?

6.5. In the library locate a good example of a gene mutation with pleiotropic effects. List the syndrome, attempting to organize the sequence of causes and effects in a manner analogous in general format to the arrangement of the data in Fig. 6.7.

6.6. What properties would you predict for a mutant that is feedback-resistant to tryptophan and, simultaneously, has a genetic block after anthranilic acid? (See Fig. 4.1 for scheme of tryptophan biosynthesis.)

6.7. Set up several theoretical models of reactions sequences and inhibitions that can account for genes in which both dominant and recessive point mutations occur.

6.8. Can you think of three different ways in which a gene mutation can produce an allele that is dominant to the wild-type allele? Outline as a theoretical series of biochemical reactions.

6.9. Some mutations are not expressed in all individuals that carry the mutant gene (incomplete penetrance). Briefly outline some reasons that might offer a biochemical basis for this phenomenon.

6.10. There is a reaction sequence, compound A to compound E, that is controlled by four genes, *1–4*: $A \xrightarrow{1} B \xrightarrow{2} C \xrightarrow{3} D \xrightarrow{4} E$. Compound E is essential for cell growth. Describe what effects mutations in each gene might have with respect to enzyme activity, accumulations, and behavior of the mutants in cross-feeding experiments and in growth tests.

Seven

Making the Messenger

Even casual observation of a higher organism tells us that all cells are not alike either in structure or in function. The cells have become "differentiated" during growth and development of the organism. All possible means of inspection tell us that each class of cells makes particular proteins and does not make others. Yet, all the cells in a fully developed metazoan are thought to be genetically identical. This basic premise has yet to be experimentally proved, but techniques are at hand that, in the near future, should permit a critical test.

In microorganisms, with their varied life cycles, certain proteins are synthesized and certain structures (for example, spores) are formed at one period, while at other times vegetative growth and cell duplication predominate. Even when viruses infect sensitive cells (see Stahl's *The Mechanics of Inheritance*), some virus-induced proteins are made immediately upon infection, while others do not appear until a later stage.

From such considerations, one is led to the proposition that *all genes* are *not expressed* at *all times* during the life of the organism. Recalling the biochemical mechanisms of protein synthesis (Chapter 3), one sees that an obvious way in which selective gene action might be achieved is through regulation of messenger RNA synthesis. Such a system could have several alternative mechanisms through which control is established. Some genes may produce many mRNA templates with great frequency, while other genes function in this capacity rarely. That is, primary gene function may be "switched on and off" at a frequency characteristic for a particular gene under a particular set of environmental conditions. Or, *equal numbers* of messenger molecules may be made along every gene, but built into each mRNA molecule may be a code, read by the protein-synthesizing machinery, that determines *how many times* the messenger will be used. Thus, a regulation hypothesis could be based on either the total number of times a specific mRNA molecule is used as a template before being destroyed, the frequency with which it is used as a protein-building template during a fixed life span, or the total number of copies of the specific messenger produced by the gene.

Current evidence indicates that mechanisms regulating gene function reside both in the control of mRNA synthesis (in control of primary gene action) and in the synthesis of proteins along the mRNA. This chapter focuses attention on experiments that show how and why regulation of mRNA synthesis is critical in gene action. In Chapter 8, other means of controlling gene function will be considered.

Watson–Crick base pairing

In its intracellular form, DNA has the double-stranded structure described in Stahl's *The Mechanics of Inheritance*. DNA acts as primer for the enzyme RNA polymerase and imparts the code, or sequence of nucleotides, to the polyribonucleotide product mRNA, which, extracted from cells, is single-stranded. For the transcription of genetic information from the double-stranded DNA molecule to the single-stranded mRNA, two general mechanisms originally were considered. (1) The mRNA is made along the outside (i.e., in the deep groove) of the DNA double-stranded helix. For example, an AT base pair could code for A in the mRNA and a TA base pair could code for U in mRNA. (2) During mRNA synthesis, the two strands of the DNA molecule separate transiently and allow specific base pairing with ribonucleotides along one of the DNA strands.

The first hypothesis requires a triple-stranded configuration with types of base pairing that are as yet unsubstantiated by laboratory experiment, while the second hypothesis utilizes Watson–Crick base pairing. The critical data, reviewed in the remainder of this chapter, are consistent with the second hypothesis, namely that one strand of the double-stranded DNA serves as template for mRNA synthesis. Since this hypothesis requires no additional *ad hoc* assumptions, it is the mechanism favored at the present time.

DNA-RNA hybrids

DNA-RNA hybrid segments are found in material extracted from cells and are formed in vitro when RNA polymerase is primed with DNA. A schematic illustration of how such a hybrid portion of a DNA molecule might be formed in vivo is shown in Fig. 7.1a. The hybrid complex is relatively insensitive to digestion by the enzymes deoxyribonuclease, which can attack free DNA, and ribonuclease, which hydrolyzes free, single-stranded RNA. The DNA-RNA hybrid can be rendered susceptible to such enzymic digestion by exposure to heat, which causes the individual strands of the complex to separate. A similar separation of the double-stranded DNA structure into two disorganized single chains also occurs upon exposure to heat.

DNA-RNA complexes also can be produced in vitro by mixing mRNA with DNA extracted from genetically identical or very closely related organisms. The complexing occurs only when the DNA is prepared in the single-stranded form (Fig. 7.1b). No complex formation

Fig. 7.1a. A portion of a DNA molecule along which mRNA has been synthesized in vivo. This hybrid segment is insusceptible to digestion with nucleases specific for DNA or for RNA.

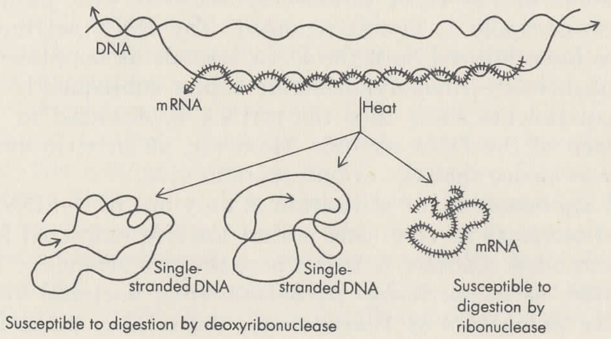
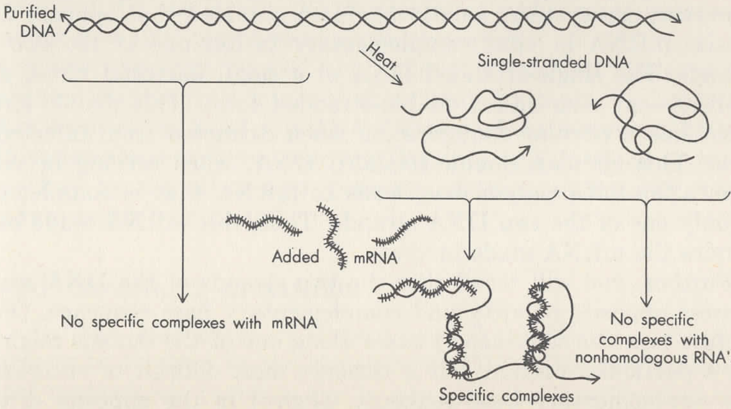

DNA
mRNA
Heat
Single-stranded DNA
Single-stranded DNA
mRNA
Susceptible to digestion by deoxyribonuclease
Susceptible to digestion by ribonuclease

Fig. 7.1b. Formation of RNA-DNA complexes between specific mRNA and single-stranded DNA extracted from the same organism. No specific complexes form between mRNA and double-stranded DNA, or between mRNA of one organism and DNA from another, nonhomologous, organism.


Purified DNA
Heat
Single-stranded DNA
Added mRNA
No specific complexes with mRNA
No specific complexes with nonhomologous RNA's
Specific complexes

is observed between mRNA and DNA of genetically dissimilar organisms, even if they have very similar over-all base compositions.

Even greater selectivity in DNA-RNA hybrid formation has been demonstrated. mRNA that has been synthesized along a particular segment of DNA will complex only with DNA from the wild-type organism. It will not complex with the DNA of a mutant whose homologous DNA segment has been deleted. This striking finding indicates that the process of DNA-RNA hybrid formation is not only organism-specific but also gene-specific.

Certainly, what is implied here is the need for identical or very similar base sequences.

Two strands, one messenger

When double-stranded DNA serves as the template in vitro, at least some of the RNA product synthesized with purified RNA polymerase normally is double-stranded. The RNA polymerase thus appears to have utilized both the DNA strands as templates, and the two complementary ribopolynucleotide copies subsequently pair. Annealing experiments show that the mRNA synthesized in vitro will bind to each of the DNA strands. However, an in vitro event is not conclusive evidence that the event occurs in vivo.

Several arguments favor utilization of only one of the DNA strands as natural template in vivo. One is that mRNA, extracted from cells, is single-stranded. Another is that where the two strands of DNA can be separated, as in particular DNA-containing bacterial viruses (see Stahl's *The Mechanics of Inheritance*), the mRNA made in the infected bacteria anneals with only one of the two strands. The mRNA has a base composition complementary to that of its DNA annealing partner and identical (except that U replaces T) with that of the second, unannealed DNA strand. No messenger is made that has affinity for the second strand, so that evidently some factor(s) dictates which DNA strand will be transcribed into mRNA. Such selectivity apparently does not normally operate during mRNA synthesis in vitro. However, one system recently has been obtained that appears to synthesize mRNA in vitro complementary to but one of the two DNA strands. The single-stranded DNA of a small bacterial virus, øX174, replicates in vivo into a double-stranded form. This double-stranded DNA has a circular configuration when extracted from infected bacteria. This circular, double-stranded DNA, when serving in vitro as primer for RNA polymerase, leads to mRNA that is complementary to only one of the two DNA strands. Thus, this mRNA made in vitro mimics the mRNA made in vivo.

Further, you will recall that the two strands of the DNA molecule possess opposite polarity and complementary base sequence. One can picture how the sequence of bases along one of the strands might code for a particular protein, but it becomes most difficult to visualize how the complementary base sequence, aligned in the opposite direction, could also code for the same protein. In spite of a versatile code, the restrictions on amino acid sequence would be severe. That the complementary base sequence of the second DNA strand does not code for a different protein is verified by direct biochemical-genetic observation: the one gene, one polypeptide chain relationship.

Finally, you will see, in the discussion of induced messenger mutations, further evidence for the suggestion that only one of the two strands of DNA is transcribed into a nucleotide sequence for a particular mRNA. If this is in fact the mechanism for mRNA synthesis, then the DNA code (nucleotide sequence) must contain information that "informs" RNA polymerase which strand to copy, that is, where to begin mRNA synthesis.

Messenger mutations

Errors that occur during mRNA synthesis appear to be rare —just as are the heritable errors, mutations, that take place during DNA replication. Mutations in mRNA are not so important in parent-to-progeny transmission, because they are not replicated and only a relatively few protein molecules out of the total number of that type in the cell may be made along the altered messenger template.

Usually, one cannot recognize small numbers of altered protein molecules in a large population of molecules: thus, mRNA mutations are not as readily detectable as are mutations in the genetic material. Only a few biological systems are known currently that offer markers sensitive enough to provide resolution at this level. One example is a colorless mutant of the flowering plant *Delphinium,* where rare single cells show color while the colorless mutation itself persists stably through the germ line. Another example of a uniquely sensitive and useful marker is the motility of bacteria. Among mutant, nonmotile bacteria that lack flagella, rare cells may be found that have transiently gained the ability to produce flagella (Fig. 7.2). They are genotypically nonmotile, and their progeny for the most part lack motility. Nevertheless, the ability to synthesize active flagella can occur as a rare, random event. In one strain, the frequency of such events was found to be about 4 × 10^{-5} per bacterium per generation.

The above examples suggest that mistakes may be made in the transcription mechanisms involved in protein synthesis. Probably the mistakes occur at the level of mRNA synthesis, for a number of altered protein molecules seem to appear together in a cell in which an event occurs. Only by the all-or-none property of color, or motility, could such a rare event be detected.

Induced messenger mutations

In *The Mechanics of Inheritance,* Stahl discusses chemical mutagenesis in relation to the structure of DNA. One mechanism of chemical mutagenesis involves the *in situ* chemical modification of a

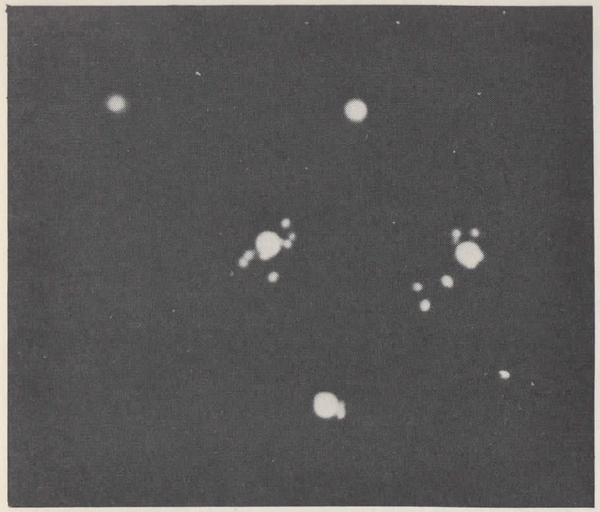


Fig. 7.2. Rare, spontaneously occurring motile bacteria in large populations of a nonmotile mutant. Parent colonies of one to several billion nonmotile bacteria were embedded in soft agar. During the growth of these colonies, a few bacteria acquired motility, although they were genotypically nonmotile. They migrated from the large parent colonies and established the small satellite colonies, also genotypically nonmotile. The number of satellite colonies per parent colony is randomly distributed. From C. Quadling and B. A. D. Stocker, *J. Gen. Microbiol., 17* (1957), 424.

base in the DNA so that nucleotide pairing on the strand opposite the chemically altered base is changed at the next round of DNA synthesis. Hydroxylamine is an example of a mutagen that causes such transition-type mutations, from a GC base pair to an AT base pair (Fig. 7.3a).

A second general mechanism of chemical mutagenesis is the incorporation, during DNA replication, of unusual nucleotides into newly synthesized strands of DNA. Once the base analogues are incorporated into DNA, there is a greatly increased probability that they will pair with the "wrong" complementary base in subsequent replications. The two compounds 2-aminopurine and 5-bromodeoxyuridine (Fig. 7.3a) are examples of chemical mutagens that are base analogues.

Another base analogue, 5-fluorouracil (5FU), is converted in vivo to the nucleoside triphosphate, and this nucleotide is then built into mRNA, replacing up to 40 per cent of the uracil in the ribopolynucleotide. Under certain experimental conditions, little or no incorporation of 5FU occurs in sRNA, ribosomal RNA, or DNA. 5FU has properties that suggest that, once incorporated into mRNA, it could occasionally attract guanine rather than the more usual complementary base, adenine. In so doing, its positioning in mRNA would exactly counteract a base substitution that previously had occurred in the DNA. This (theoretical) process is diagrammed in Fig. 7.3b.

This figure shows a segment of the DNA molecule associated with the synthesis of a segment of an mRNA molecule. During protein synthesis, the mRNA serves as template in attracting particular amino acid–charged sRNA molecules. When the wild-type configuration in the DNA molecule is CG and the mRNA contains C (7.3b, upper left),

Fig. 7.3a. Predominant mutagenic effects of three compounds on T4 bacteriophage. For detailed information on chemical mutagenesis, see E. Freese, in *Molecular Genetics*, Part I, J. H. Taylor, ed. (Academic, 1963).

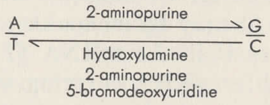
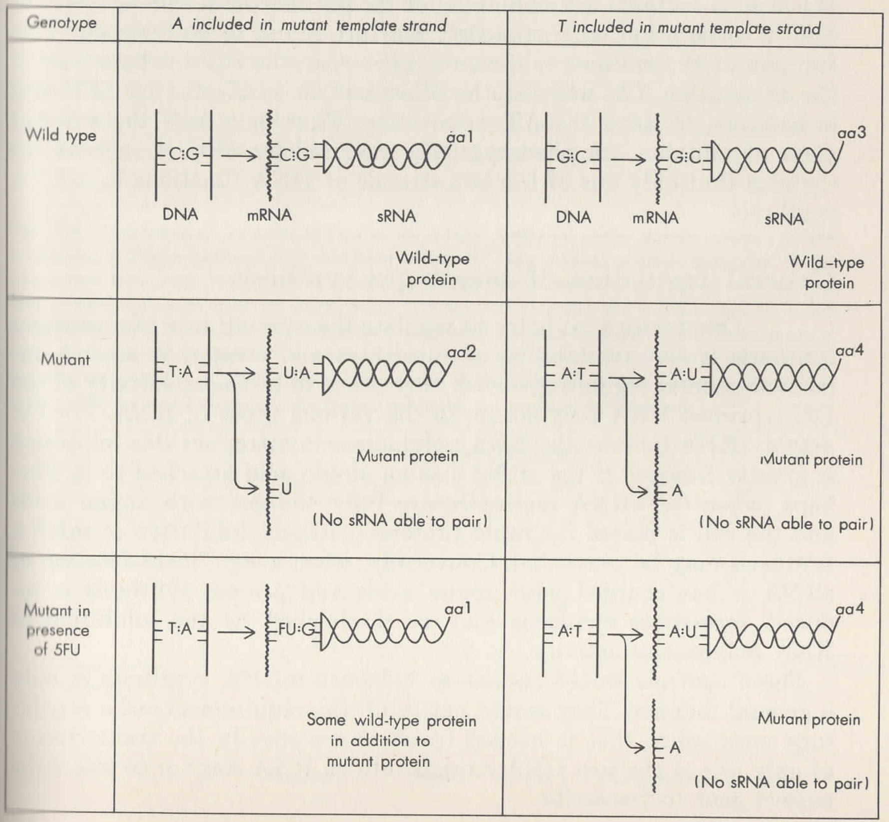

A
T
2-aminopurine
G
C
Hydroxylamine
2-aminopurine
5-bromodeoxyuridine

Fig. 7.3b. Theoretical picture of predominant mutagenic effects of 5FU in mRNA synthesis. A group of mutants is selected that shows an increased rate of reversion to wild type in the presence of 2-aminopurine but no appreciable increase in the presence of 5-bromodeoxyuridine or after treatment with hydroxylamine. Mutant genes of this class are assumed to carry an adenine-thymine (AT) base pair in place of the guanine-cytosine (GC) pair characteristic of wild type. The AT base pair may be arranged in the DNA either so the strand carrying A serves as mRNA template or so the strand carrying T serves as mRNA template. If A is included in the mutant template, incorporation of 5FU into mRNA allows occasional pairing with G in sRNA and partially phenotypically reverses the effect of the mutation in the DNA. If T is included in the mutant template strand, FU has no effect. Approximately half of the total number of mutants in the group fall into each category: FU-reversible and FU-nonreversible.


Genotype
A included in mutant template strand
T included in mutant template strand
Wild type
C:G
C:G
aa1
DNA
mRNA
sRNA
Wild-type protein
G:C
G:C
aa3
DNA
mRNA
sRNA
Wild-type protein
Mutant
T:A
U:A
aa2
U
Mutant protein
(No sRNA able to pair)
A:T
A:U
aa4
A
Mutant protein
(No sRNA able to pair)
Mutant in presence of 5FU
T:A
FU:G
aa1
Some wild-type protein in addition to mutant protein
A:T
A:U
aa4
A
Mutant protein
(No sRNA able to pair)

an sRNA molecule containing code letter G becomes attached at this site. This fixes the location of amino acid 1. But if, as a consequence of mutation, the CG base pair has been replaced by a TA base pair (7.3b, middle left), the mRNA molecule containing U as part of its coding unit may attract an sRNA molecule with A as code letter and make mutant protein, or it may be incapable of attracting any sRNA. However, if 5FU replaces U in the mRNA (7.3b, lower left), pairing will still be possible with an sRNA carrying amino acid 1, the amino acid present in the wild-type protein.

Note the consequences when the wild-type CG base pair is reversed, so that the functional base, in mRNA biosynthesis, is C (7.3b, right). The mutant mRNA copied from T of the mutant DNA now contains A. 5FU does not replace A and so is not incorporated at this position in the mRNA; thus, the mutational alteration in the DNA is not counteracted.

According to the rules of chemical mutagenesis, a group of mutants can be selected that represent changes from GC to AT base pairs in the DNA. These are mutants that are caused to mutate back to wild type by incorporation of 2-aminopurine but are not appreciably back-mutated by hydroxylamine or by 5-bromodeoxyuridine (Fig. 7.3a). When such mutants are examined for the phenotypic reversing effect of 5FU incorporation, *approximately half* are found to be reversed. This fraction of the mutants is assumed to contain the mutant base pair in the orientation TA, whereas the other half is unaffected by 5FU and is assumed to have the AT orientation. That only half the selected class of mutants are phenotypically reversed by 5FU is support for the idea that only one of the two strands of DNA functions in mRNA synthesis.

General regulation of messenger synthesis

One factor that helps to regulate the over-all rate of messenger synthesis is the availability of substrates and energy. A second important general regulating device may prove to be the sensitivity of the DNA-primed RNA polymerase to the various types of RNA. For example, sRNA inhibits the RNA polymerase in vitro, but this inhibition is greatly lessened if the sRNA has an amino acid attached to it. Perhaps, when the sRNA molecules are fully charged with amino acids and the cell is poised for rapid protein synthesis, inhibition of mRNA synthesis may be prevented. Conversely, when a significant fraction of sRNA is not charged with amino acids and protein synthesis is reduced, messenger synthesis may be diminished by the inhibition of RNA polymerase activity.

These controls would appear to influence mRNA synthesis in only a general manner. They would not fulfill the requirements of a regulatory mechanism that is needed to effect specifically the transcription of only *one* of the two DNA strands into an RNA copy or to select the correct gene to transcribe.

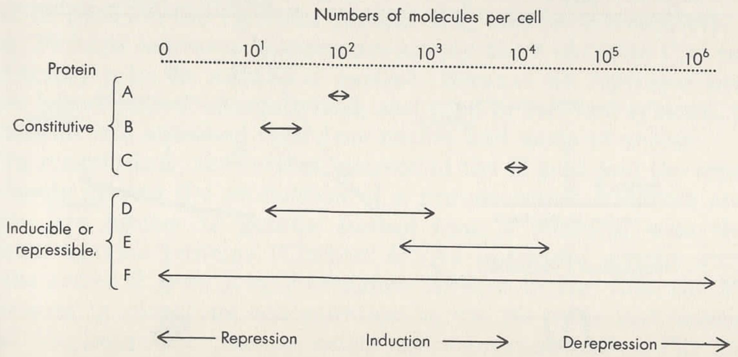
Induction and repression

In higher organisms, most of the protein in one type of cell may constitute a single protein species that is undetectable in other types of cells from the same organism. Also, a particular protein may be made in abundance at one stage during the development of an organism but appear only in very low quantity, or be undetectable, in other stages. The differential expression of genic activities in higher organisms is discussed by Clement Markert in *Developmental Genetics,* a companion volume in this series, and will not be discussed extensively here. Instead, we shall focus our attention on well-analyzed phenomena in microorganisms that give some clues to mechanisms regulating gene action.

In microorganisms, a single protein species may be absent under some conditions of cultivation and yet constitute several per cent of the total protein of the cell when the microbes are cultured under a different set of conditions. Figure 7.4 illustrates how widely the levels of some proteins may vary. Also note from the figure that there are other proteins that are formed in relatively constant amounts regardless of different conditions of culture.

Proteins that fluctuate widely in amounts as a function of changes in the environment of the cell are called inducible proteins, if their level increases following the addition of a metabolite, or repressible proteins, if their level decreases after addition of a metabolite. If the level of a particular protein increases after removing a specific metabolite, the phenomenon is termed derepression.

Fig. 7.4. Numbers of protein molecules in single bacterial cells grown under various conditions. A single bacterial cell can have about 10^7 (ten million) protein molecules. If the generation time (time for complete doubling by vegetative reproduction) is 30 minutes, 10^7 new protein molecules must be made every half-hour to maintain the protein content in the progeny bacteria. Some proteins are present in greater abundance than others. Some proteins vary in concentration but little when the bacteria are grown in different environments (constitutive proteins). Other proteins show wide fluctuations when the bacteria are placed under different environmental conditions (inducible, or repressible, proteins).



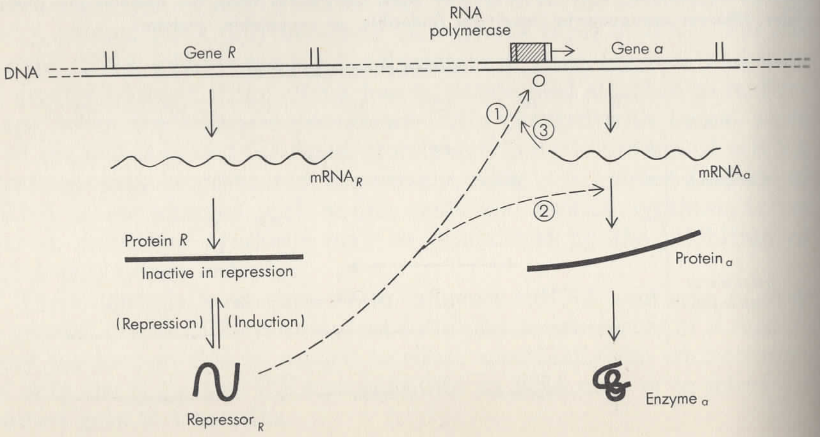
Proteins that do not fluctuate appreciably in amounts under the selected set of growth conditions are termed constitutive proteins. It is possible that all protein levels are regulated by metabolites, but that only the synthesis of certain proteins can be regulated by the addition of metabolites to the external environment. The intracellular environment can be altered in living organisms only within relatively narrow limits as compared with the wide variations possible in the surrounding milieu.

Biochemical and genetical studies of numerous gene regulatory systems have implicated four key components in the control of gene function (Fig. 7.5). They are: (1) a gene for specification of a repressor protein, (2) the repressor protein, (3) a gene or gene cluster whose activity is regulated by the repressor protein, and (4) an operator region, at one end of the gene, or gene cluster, that is sensitive to the repressor protein or that produces a repressor-sensitive molecule. It should be emphasized that the model shown in Fig 7.5 is, in broad outline, substantiated by experiment; however, gaps in knowledge are considerable and many of the details of the model are still speculative.

Fig. 7.5. A general model for induction and repression of specific protein biosynthesis. The figure shows some of the major components that are presumed to take part in the regulation of genic activity and some possible sites of interaction by the components.

A gene, *R*, is responsible for the synthesis of a protein that can exist in two conformations. In one conformation the protein is inactive as a repressor. In the second configuration (bottom) the protein has affinity for a chromosomal location, O (step 1). When the repressor protein is at O, RNA polymerase is unable to synthesize mRNA along gene *a*. When repressor protein is absent from O, RNA polymerase initiates synthesis of mRNA. A modification of this general scheme is diagrammed as reactions 2 and 3. In this modified model, the repressor acts at the level of mRNA, causing it to remain complexed with the structural gene and to halt further mRNA synthesis at this place on the chromosome.

The amount of the repressor available for inhibition of mRNA synthesis depends upon the rate of synthesis of the repressor protein, upon the repressor's affinity for O, and upon the presence of molecules of small molecular weight that influence the interconversion of the two forms of the repressor. In induction, the metabolite maintains the repressor molecule in a form that is inactive as a repressor. In repression, the metabolite enhances the conversion of repressor molecules from the inactive form to a conformation that is active as repressor.



The R *gene and the repressor*

The repressor molecule, currently thought to be a protein, is pictured as existing in two alternative and specific conformations, one active in repression and the other inactive. The primary structure of the repressor protein presumably dictates to a large extent which of these two forms will be favored. Because the primary structure of the repressor molecule is subject to alteration by mutations in gene *R*, the proportion of the protein that is active as repressor is under the control of gene *R*.

Metabolites of small molecular weight are visualized as affecting the ratio of the two protein forms by complexing with one of them. The metabolite that increases the concentration of the active repressor increases enzymic repression and is called a corepressor. The metabolite that increases the proportion of the inactive repressor decreases repression and is known as an inducer (Fig. 7.5).

Some of the mutations in the *R* gene, especially a deletion of the gene, can be pictured best as a complete shutting off of the production of repressor protein. Other mutations in the *R* gene that also lead to constitutive enzyme production may elicit formation of a repressor protein with a greatly decreased ability to attain the active configuration essential for interaction with the operator site. The mutationally altered repressor protein may fail to bind the corepressor readily or may not undergo the proper conformational change to active repressor after the binding of corepressor. Finally, other mutations in *R* genes lead to a hyperrepressed condition in which enzyme production at gene *a* is reduced below normal. Such hyperrepressed mutations are postulated to act by eliciting the production of active repressor protein that has a very high affinity for the sensitive operator site or that more readily assumes the active repressor configuration in the presence of metabolites normally found in the cell.

The characterization of a gene as a repressor gene does not imply that its product, the repressor protein, may not have a catalytic function. Perhaps repressor proteins are simply those enzymes that possess additional roles in metabolic control. Because no repressor proteins have been isolated, characterized, and used in cell-free systems, many mysteries still surround their true nature and mode of action.

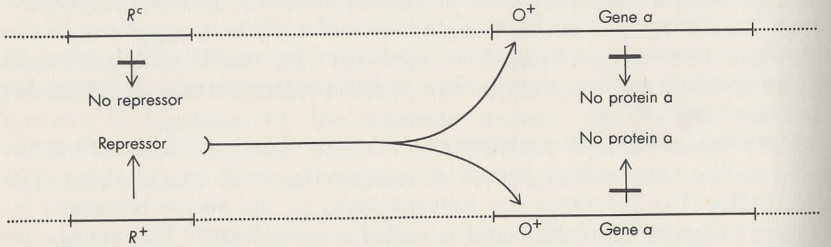
To recapitulate, the current concept of the *R* gene and the repressor molecule depicts the production of a protein whose structure and activity are subject to genetic control that is identical with that of normal enzyme proteins (Chapter 4). An important genetic attribute of the active *R* gene is its dominance. As seen in Fig. 7.6a, the *R* gene functions in the *trans* configuration in the *cis-trans* test, suggesting that the repressor protein exists extrachromosomally. The unique feature of the *R* gene is its dictation of the structure of a protein that

Fig. 7.6. Dominance relationships of *R* and O genes.

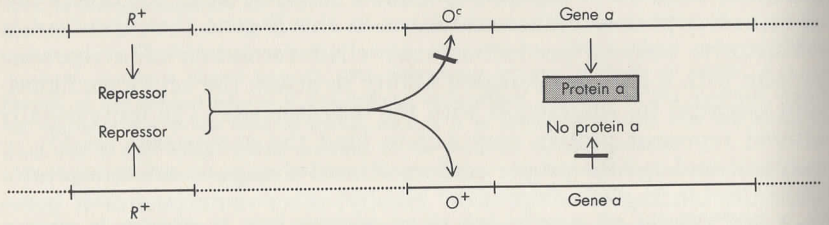
Symbols

R^+ = wild-type repressor gene
R^c = constitutive mutant
O^+ = wild-type operator region
O^c = constitutive mutant
O^0 = operator zero (operator negative) mutant

(*a*) Production of repressor is dominant in the *trans* configuration. When two sets of genes are in the same cell, the functioning of a single repressor gene dictates the formation of an extrachromosomal product capable of affecting (repressing) any susceptible operator region.



(*b*) Operator mutations (O^c *and* O^0) do not affect the functioning of genes on the second (*trans*) element; they affect only those genes *cis* to operator. In operator constitutive mutants (O^c), protein production continues in the presence of repressor.



can interact in a site-specific manner with the operator or with a product of the operator, thus possibly influencing the production of mRNA by the structural gene adjacent to the site of interaction.

The operator

Operators are believed to serve two functions. They are starting places for mRNA synthesis along one of the strands of the DNA, and either they or their product can interact with a specific repressor molecule. Mutations in an operator, then, might be expected to affect one or both of these functions.

The operator mutations that shut off or greatly slow down mRNA synthesis are termed O^0 (operator zero) mutations. Mutations in an operator region that reduce mRNA synthesis can be due to one of four causes: (1) loss of ability to attract RNA polymerase, which may occur, for example, when the operator is deleted, (2) increased affinity of the operator or its product for the specific repressor molecule, (3) affinity for a repressor molecule that has little or no affinity for the wild-type operator or its product, or (4) an alteration in nucleotide sequence that affects the translation of the messenger into protein (see Chapter 8).

Mutations in operator that lead to increased or even to completely constitutive enzyme production are termed O^c (operator constitutive) mutations. Such mutations presumably lessen or eliminate completely the affinity of operator or its product for the specific repressor molecule.

Noteworthy is the fact that O^0 and O^c mutations affect only those genes located *cis* to the mutant operator (that is, only those genes contained on the same DNA molecule). Failure of operator mutations to function in the *trans* configuration is taken as evidence that the operator serves as a receptor site and does not exert its effect through the production of an independent cytoplasmic product (Fig. 7.6). Furthermore, and in support of this view, it appears that constitutive protein synthesis continues in a cell that contains an O^c mutation on one chromosome and a normal repressor-susceptible operator on the homologous chromosome.

It is not known whether operator is separate from the structural gene whose function it controls or whether it extends into the adjacent structural gene. Thus, it is possible that some operator mutations affect both the primary structure of a protein and its rate of synthesis as well. Comparison of the primary structure of a constitutively produced protein with its inducible (or repressible) counterpart is necessary to resolve this point.

The operon

In the preceding sections, we have considered, as a theoretical example of an operator, one that affects the production of only one enzyme, protein a. However, in many bacterial systems, several genes that code the structure of several different proteins may be controlled by a single operator. Such genes are closely linked on the chromosome. The segment of genetic material that bears the functionally coordinated multigene complex of operator and structural gene(s) is called an operon.

One example of an operon is the cluster of genes that determine the structure of the enzymes necessary for histidine biosynthesis (Fig. 4.5). Although each gene codes for an individual protein, the genes function in a coordinated fashion. Wild-type bacteria make sufficient histidine for maximal repression of enzyme levels to a uniform value. As a result, the addition of histidine to the medium has no effect on these basal enzyme levels. The basal levels are arbitrarily set at one unit for each enzyme. Histidine-requiring mutants, grown on histidine, usually produce only one unit of each of those enzymes that are still functional in the pathway. However, when the intracellular concentration of histidine is limited, the enzymes are produced in increased and equal quantities. The amount of increase is set by the exact cultural conditions; when, for example, they are adjusted so that one enzyme increases tenfold, each of the other enzymes increases by precisely the same amount. The important point is that the changes in enzyme activities are coordinate (Fig. 7.7). Glutamic dehydrogenase, controlled by a gene outside the histidine operon, is not affected by histidine in this manner (7.7, right).

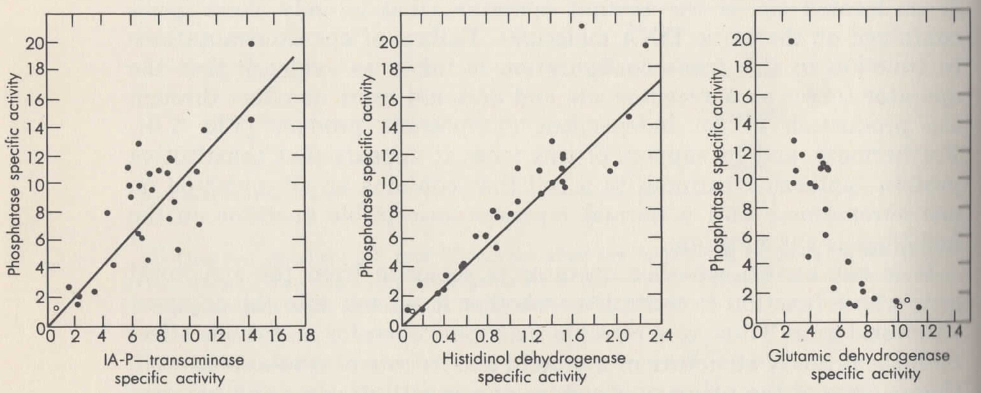


Fig. 7.7. Coordinate repression of enzymes involved in histidine biosynthesis. The specific activity of L-histidinol phosphate ester phosphatase (gene *B*) plotted against the specific activities of imidazoleacetol phosphate ester transaminase (gene C) at the left, L-histidinol dehydrogenase (gene *D*) in the center, and glutamic dehydrogenase (not controlled by the histidine operon) at the right. Each point represents assays for the specified enzymes on a single extract. The extracts were prepared from bacteria grown under different cultural conditions allowing varying degrees of derepression. Extracts from wild-type bacteria are symbolized by open circles, extracts from mutants by solid dots. Redrawn from B. N. Ames and B. Garry, *Proc. Nat. Acad. Sci. U.S., 45* (1959), 1453.

Thus, the operator regulates the amounts of enzyme proteins synthesized by the member genes of the operon. This control is coordinate, in that each of the member genes is affected to a similar extent. In the next section, we shall see that mutations in the operator region, or in the *R* gene, can affect enzyme levels; but, again, these effects are always precisely the same for each gene in the operon—they are precisely coordinate.

Some situations are known in which several genes respond to the same repressor but the genes are not linked closely with one another; each gene or cluster of genes, together with its own operator, appears to constitute a separate operon. The operators have the common property of sensitivity to a single repressor (produced by a single *R* gene), but because they can have different affinities for the repressor, they must not be identical in structure. Consequently, the genes in each

operon are differentially affected by conditions of derepression or by *R* gene mutation. While all enzyme levels may increase, they do not increase in precise, coordinate fashion.

One example will suffice to illustrate the types of genetic evidence upon which the operon concept rests. Diagrammed below the histidine operon in Fig. 4.5 are some mutations that have been demonstrated by recombination tests to extend over the regions covered by the solid lines. It has been noted with most of these mutations that enzymes controlled by genes located outside the mutated region approach the normal level and are influenced by the presence or absence of histidine. However, the mutations that extend to the "right" end of the operon (mutations *203, 63, 538, 57,* and *644*) completely shut off or enormously reduce all enzyme production characteristic of this operon. This is illustrated more clearly for mutant *203* in Fig. 7.8a-b. In mutant *203*, the histidine operator seems to be deleted and the structural genes in the operon are under the control of a second operator that is itself repressed (see Fig. 7.8d).

Some secondary mutants may be obtained in which part of the histidine operon of *203* has been duplicated, as shown in Fig. 7.8c. Although the duplicated genes are structurally identical with those already present in the chromosome, the duplicate copies function, for they are under the control of a different operator region (wavy line) now attached to the critical portion of the operon. Consequently, the enzyme level is no longer influenced by the presence of histidine.

Another kind of secondary mutant from strain *203* has its genes "turned on." Some of these represent further deletions of genetic material at the "right" end of the operon. This suggests (Fig. 7.8d) that further deletion (called *1300*, etc.) has eliminated the substitute operator and that the genes are now under the control of a new operator region. While the genes that still remain intact (genes *E, I, F, A, H, B, C, D*) function coordinately, enzyme levels are no longer influenced by the presence or absence of histidine. Histidine, the corepressor, which normally interacts with a repressor molecule specific for the histidine operon, exerts no effect on the repressor molecule specific for this new operator.

An O^c mutation also has been obtained in a strain that has an intact gene *G*. The histidine enzymes are coordinately raised about tenfold over the normal basal level; however, unlike the mutants described above, this mutant can be further derepressed another two- to threefold in the absence of histidine. This O^c mutation maps to the "right" of all other known single-site (point) mutations, next to gene *G* in the histidine operon (Fig. 4.5).

Finally, mutations in one gene that is not closely linked to the histidine operon are known to affect specifically enzyme production by the member genes of the histidine operon. Such mutations, which are presumed to be alterations in the *R* gene, are constitutive mutations. The levels of the histidine enzymes are coordinately raised over the normal level in the presence of histidine.

Thus, the operator is the crucial site for the regulation of the synthesis of each of the proteins coded by the member genes of the operon.

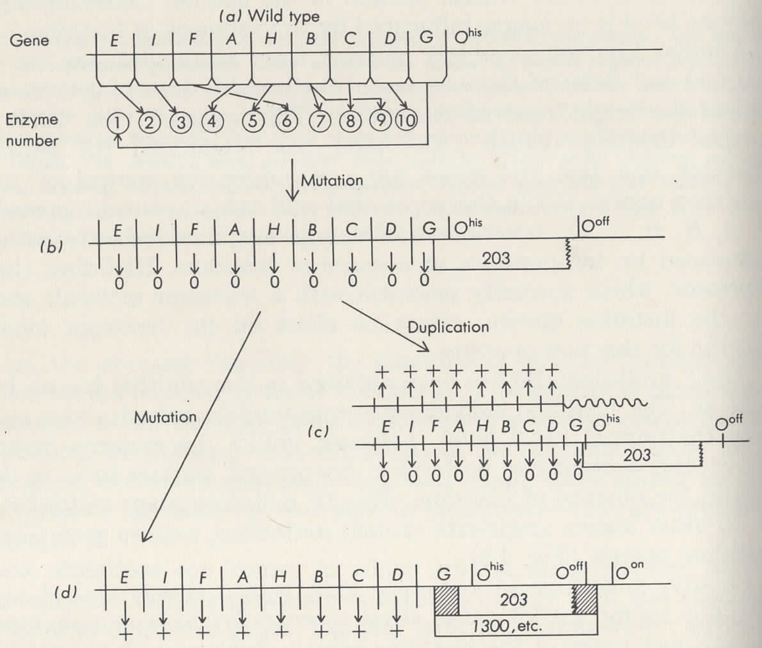
Fig. 7.8. Interpretation of genetical changes that simultaneously influence enzyme production by a coordinated series of eight linked genes.

(*a*) The histidine operon of wild-type bacteria, showing the arrangement of the genes on the chromosome (genes *A* through *I*) and the number of the biochemical reaction with which each gene is concerned.

(*b*) In a particular mutant (mutant isolate 203), the terminal portion of gene G and some nearby genetic material outside the histidine operon have been deleted. Although genes *E* through *D* are still present in intact form, they do not function. For example, the functioning of gene *D* gives rise to L-histidinol dehydrogenase, necessary for the conversion of L-histidinol to L-histidine. But mutant *hisG203*, unlike other G mutants, fails to grow on L-histidinol and lacks L-histidinol dehydrogenase activity in extracts.

(c) Through secondary mutation, the genes *E* through *D* of the histidine operon, inactive in mutant *his203*, are again active. The rare secondary mutant is detected by its ability to grow on L-histidinol. It produces the histidine enzymes 2 through 9 coordinately and constitutively. It is unstable and loses the extra genetic element, segregating off the original 203 mutant.

(*d*) A different type of secondary mutation also restores ability to grow on L-histidinol. Some of these mutations are further deletions involving portions of gene G (shaded areas) that again turn on enzyme production by the remaining intact genes *E* through *D*. Enzyme production is coordinate and constitutive.



These and other studies with microbial systems substantiate the utility of the model for investigation into genetic fine structure and its function. Certainly they make clear that genetic material is organized into specific, highly oriented regions whose proper function entails a considerable degree of interplay and interdependence at the genic and extragenic levels.

The unique mRNA molecule

The amount of a particular mRNA present in cells can be measured in two ways. One technique is illustrated in Fig. 7.9. By differentially labeling the RNA in two bacterial strains, one of which is competent genetically and is induced to make a specific type of protein and the other of which is not, a fraction of RNA resembling mRNA can be found that is unique to the strain synthesizing the specific protein. This distinct RNA fraction is readily detectable, although it constitutes only a small percentage of the total RNA synthesized. Its presence indicates that in the incompetent strain the first effect of the repression of enzyme formation has been to reduce production or eliminate a specific type of mRNA. Other experiments show that uniquely labeled mRNA fractions have the ability to form specific hybrid molecules with only a small section of the bacterial DNA, namely, that section containing the appropriate homologous genes.

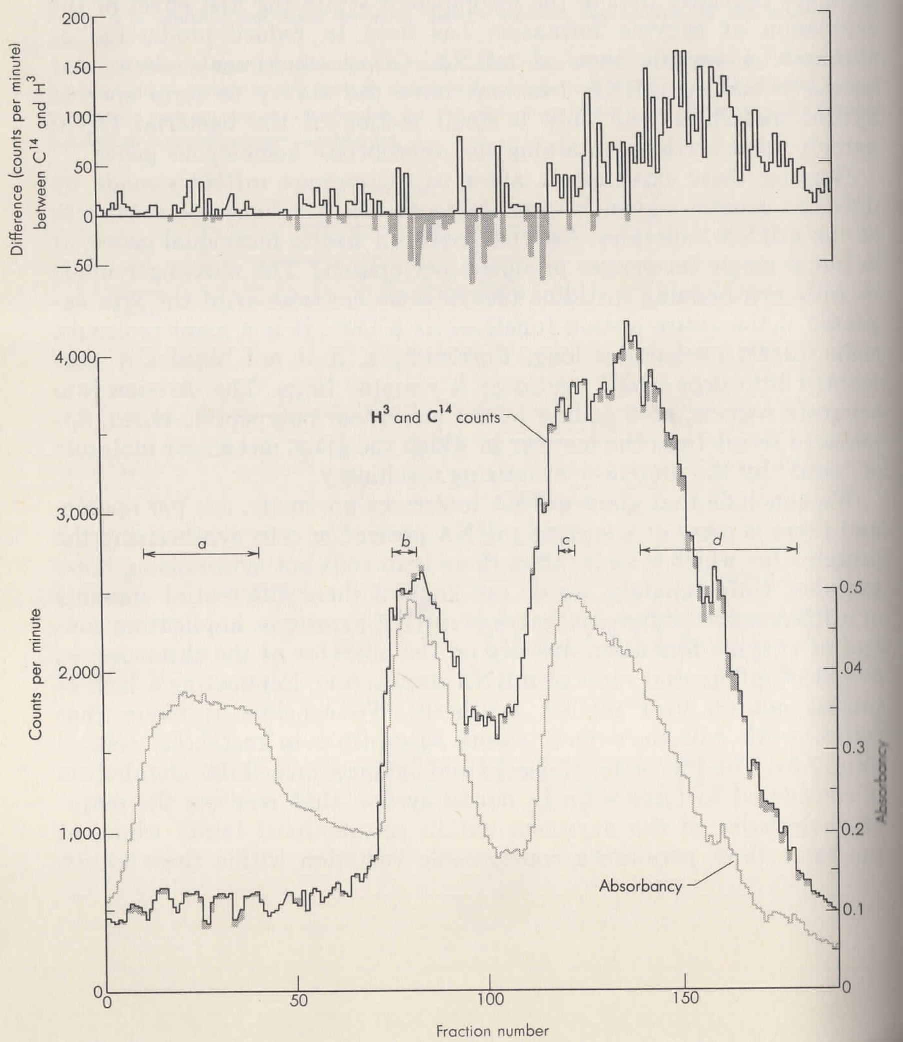
Because these experiments allow us to compare mRNA's made by different genetic regions, we can begin to ask questions about the size of the mRNA molecules. Are they equal in size to individual genes, or is but a single messenger produced per operon? The messenger made in cells synthesizing histidine-biosynthetic enzymes is of the size expected if the entire operon functions as a unit. It is a giant molecule, some 13,000 nucleotides long. Furthermore, it is not physically segmented into gene-length sections; it remains large. The division into separate regions, each coding for an individual polypeptide chain, appears to result from the manner in which the giant messenger molecule is "read" by the protein-synthesizing machinery.

We conclude that giant mRNA molecules are made, one per operon, and there is more of a specific mRNA present in cells synthesizing the proteins for which it codes than there is in cells not synthesizing these proteins. Unfortunately, we do not know if these differential amounts of mRNA reflect differential rates of mRNA synthesis, implicating control of enzyme formation directly on the operator of the chromosome, or reflect differential rates of mRNA destruction, implicating a lack of precise control over mRNA synthesis. We consider it likely that future work will show both points susceptible to metabolic control (Fig. 7.5). This hierarchy of checks and balances on cellular metabolism is considered to form a finely poised system that restricts the major characteristics of the organism within certain fixed limits while, at the same time, permitting considerable variation within these limits.

Fig. 7.9. Detection of mRNA for the histidine operon. Mutant *hisl52*, deleted in six of the nine histidine genes (see Fig. 7.8) is grown with a two-minute pulse of H^3-uridine, while a constitutive mutant that forms high levels of enzyme is grown with a similar pulse of C^{14}-uridine. The cells are harvested, mixed, and washed, and the RNA is isolated by column chromatography.

The lower portion of the figure shows the measure of nucleic acid in each fraction (absorbancy), tritium (H^3), and radioactive carbon (C^{14}). The first peak (at *a*) is DNA; the second peak (at *b*) is 16s RNA; the third peak (at c) is 23s RNA. At least some of the RNA in these two peaks is ribosomal RNA. The region to the right of the 23s RNA (at *d*) is considered to be larger than 23s RNA (that is, larger than one million molecular weight). In this heterogeneous, rapidly labeled RNA, an RNA fraction is found that is unique to the mutant synthesizing the histidine enzymes. This is not clear from the lower portion of the figure but is strikingly demonstrated in the upper portion of the figure (ratio of C^{14} to H^3 counts). It represents about one per cent of the pulse-labeled RNA that is adsorbed to the column. When the material at *d* was pooled and sedimented in a sucrose density gradient, the differentially labeled material had a sedimentation coefficient of approximately 38s, indicating a molecular weight of about four million.

No significant difference in profile was observed between RNA from a repressed strain and RNA from the deletion mutant, *hisl52*. Therefore, there is more of a certain kind of mRNA present in derepressed bacteria than in repressed bacteria. Redrawn from R. G. Martin, *Cold Spring Harbor Symp. Quant. Biol.*, 28 (1963), 357.



References

Ames, B. N., and R. A. Martin, "Biochemical Aspects of Genetics: The Operon," *Ann. Rev. Biochem.*, *33* (1964) 235. An excellent recent summary.

Bonner, D. M., ed., *Control Mechanisms in Cellular Processes*. New York: The Ronald Press Company, 1961.

Cold Spring Harbor Symp. Quant. Biol., *26* (1961) and *28* (1963). Symposium volumes containing reviews and original contributions toward analysis of control mechanisms with emphasis on systems in microorganisms.

Jacob, F., and J. Monod, "Genetic Regulatory Mechanisms in the Synthesis of Proteins," *J. Mol. Biol.*, *3* (1961), 318. A creative review that represents the clearest presentation of the operon-repressor gene concept and serves as stimulus in this area of research.

Vogel, H. J., V. Bryson, and J. O. Lampen, eds., *Informational Macromolecules*. New York: Academic Press, Inc., 1963.

Questions

7.1. Draw the Watson–Crick base pairs (*a*) in DNA, (*b*) in an RNA-DNA hybrid, (*c*) in RNA.

7.2. Describe some ways in which mutation can affect functioning of a gene, gene *A*, without affecting the structure of the protein controlled by gene *A*.

7.3. Explain why operator mutations are effective only on genes *cis* to them. Explain why mutations in repressor genes are effective both on genes of the same genetic element and on genes in other genetic elements.

7.4. Suppose that tryptophan were completely effective in repressing the genes of tryptophan biosynthesis. (Note that repression is ordinarily incomplete.) What phenotypes would you expect to find in mutants containing transpositions whereby the genes for histidine biosynthesis had become transposed intact into the tryptophan operon?

7.5. Describe the selective advantages that you believe feedback inhibition and repression might offer organisms.

7.6. Outline the system of genetic control of β-galactosidase synthesis in *Escherichia coli* as described in books or review papers in the library.

7.7. Pick out one method, described in this chapter or in earlier chapters, that you think might be useful in analysis of growth and development in higher organisms and tell why.

7.8. Elaborate at least two ways an organism could become constitutive for an enzyme through a single gene mutation.

7.9. If only one DNA strand is used in making the messenger, why do you think the organism bothers to make the second strand?

7.10. What are "messenger mutations" and how do they occur?

Eight

Regulation on the Ribosome

In Chapter 7, the importance of the transcription step from the DNA code to the mRNA complementary copy was examined from two points of view: the accuracy of the transcription process and the role the rate of transcription may play in regulation of gene function. This chapter will examine how other components of the protein-synthesizing machinery affect the translation of the code in mRNA into the amino acid sequence of protein and how the messenger code is interpreted to bring about the expression of the genetic information present in the DNA. We shall see that mutations that affect the protein-synthesizing machinery can also modify the reading of the genetic code. We shall examine also the possible roles that the protein-synthesizing system itself may play in cellular mechanisms that regulate gene action.

Genes without proteins

The fact that many genes have been found to determine protein primary structure does not necessarily mean that *all* genes act in this capacity. Indeed, there is good evidence that some genes may serve as templates for the formation of sRNA and ribosomal RNA.

We know that both sRNA and ribosomal RNA can form specific complexes with DNA of the same organism and

that the formation of such hybrid complexes is a major criterion of sequence complementarity. The complex of ribosomal RNA with DNA is insensitive to nuclease digestion, and the base composition of the RNA component differs from that of mRNA. Only a small portion, about 0.2 per cent, of bacterial DNA forms specific complexes with ribosomal RNA. The conclusion is that DNA contains localized sequences of nucleotides that are complementary to ribosomal and to sRNA. As discussed in Chapter 3, the current view is that these few genes serve as templates for the synthesis of RNA molecules that have functions other than a messenger function and that such RNA molecules are not involved as *templates* for protein synthesis.

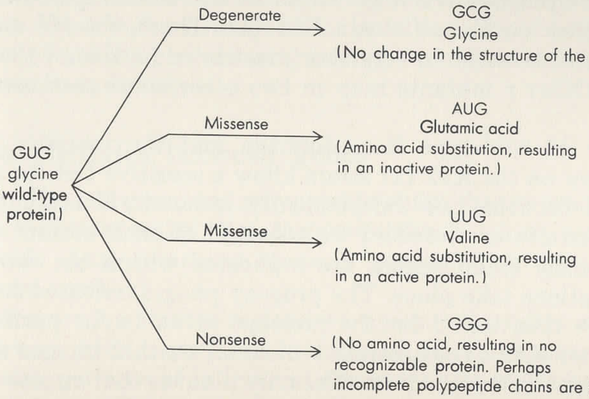
Mutations and the code—some jargon

For the moment, we can disregard the precise nature of the genetic code and, instead, examine some general ways in which mRNA coding units might be altered. Figure 8.1 shows some possibilities.

The code is considered to be degenerate (top arrow in figure), which means that there is more than one coding unit for a single amino acid. If a coding unit, for example GUG, is changed to another coding unit, for example GCG, that codes for the same amino acid, then no amino acid substitution will be detected in the protein. Under these conditions, the mutation giving rise to an altered mRNA might remain unobserved in the final phenotype.

On the other hand, the coding unit may be changed so that it now represents a coding unit for a different amino acid (missense) or for no amino acid at all (nonsense). These possibilities already have been diagrammed in Fig. 7.3. An mRNA molecule whose nucleotide sequence is a transcription of a nonsense coding unit of DNA is assumed to be

Fig. 8.1. Some possible phenotypes resulting from mutational alterations in one letter of a hypothetical coding unit.


GUG
glycine
(wild-type protein)
Degenerate
GCG
Glycine
(No change in the structure of the protein.)
Missense
AUG
Glutamic acid
(Amino acid substitution, resulting in an inactive protein.)
Missense
UUG
Valine
(Amino acid substitution, resulting in an active protein.)
Nonsense
GGG
(No amino acid, resulting in no recognizable protein. Perhaps incomplete polypeptide chains are formed.)

unable to base-pair with any sRNA molecules charged with amino acids. A missense coding unit, on the other hand, would result in the formation of an mRNA molecule that would base-pair with an sRNA molecule able to recognize the new coding unit. Obviously, quite often this sRNA molecule will carry a "wrong" amino acid, and in this way, an amino acid substitution in the polypeptide chain can occur.

The amino acid substitution caused by a missense mutation may be such that a mutant phenotype is detected readily. Conversely, the mutation may not be detected if the incorrect amino acid that has been inserted in the polypeptide chain does not appreciably change the conformation and subsequent function of the protein product. In fact, a missense mutant phenotype might be recognizable only by examination of the primary structure of the protein, not by examination of other properties of the protein or of the gross cellular phenotype.

rII: Real genes and hypothetical polypeptide chains

Genetics is such an interesting and provocative science because very often models have been constructed solely by inference from genetic data, only to be verified later by direct cytological, biochemical, or other observations. This kind of educated conjecture has been employed in studies of the *rII* region of the T4 bacteriophage chromosome. Genetic analysis of this region is a most elegant example of this approach to solving the problem of gene-protein relationships. Our picture of the fine structure of the *rII* region is based wholly on genetic experiments, and the polypeptide chains whose structures are most definitely thought to be determined by this region still remain to be found among the myriads of proteins in the living cell! It is both for convenience and as a tribute to the ultimate reliability of competent and thorough genetic analysis that the *rII* region is introduced in this chapter. The use of this system for discussion of the fine structure of the gene can be seen also in Stahl's *The Mechanics of Inheritance.*

The *r* (rapid lysis) mutants of T4 bacteriophage are detected by their plaque morphology when grown on *E. coli* strain B. Some *r* mutants fail to grow on *E. coli* strain K12 (λ). These, the *rII* mutants, map in genetic tests within a very short, restricted portion of the phage chromosome. Other *r* mutants map in two other short sections of the chromosome.

The ability of wild-type bacteriophage and the inability of *rII* mutants to grow on the K12 (λ) strain allow a sensitive test for genetic recombination between two independently isolated *rII* mutants. The sensitive B bacteria are infected with several of each mutant phage; the mutant phage chromosomes are replicated within the same cell, and recombinations take place. The progeny phages, released from the B bacteria, are then tested for the presence of wild-type particles by plating them on K12 (λ) bacteria; billions of mutant phages may be tested on a single dish of medium. The only plaques that appear will be due to spontaneous mutations to the wild-type phenotype (evaluated by control platings of each type of mutant bacteriophage grown separately where genetic recombination cannot occur) and to wild types that arise from genetic recombination. The number of recombinants thus is determined. To determine the *frequency* with which recombinant particles appear in the progeny, the progeny phages also are assayed on the sensitive B bacteria to determine the total number of particles (mutant plus wild type) present in the suspension.

The *rII* mutants are characterized, then, by a genetic block that is detected by their inability to grow in the bacterial host K12 (λ). The *cis-trans* test for complementation can be applied to the *rII* system by mixed infection of *E. coli* K12 (λ) with different *rII* mutants. The test is diagrammed in Fig. 5.1e. If K12 (λ) bacteria are mixedly infected with wild type and an *rII* single or *rII* double mutant (the *cis* configuration), the cells lyse, liberating both types of phage. The *rII* mutation is recessive. The *trans* configuration of the test is just as readily performed by mixedly infecting a culture of the K12 (λ) bacteria with a pair of *rII* mutants. If the two mutations are in two different genes, or cistrons, they complement each other, providing all the functions necessary for phage growth in K12 (λ). When the bacteria lyse, they all liberate the two mutant phage types, and a few release rare, recombinant phages. In contrast, if the two mutations affect the same function, they fail to complement each other, phage growth does not ensue, and no phage particles are released.

Mutants of the *rII* phenotype fall into two classes, A and B, on the basis of the *cis-trans* complementation test. All A mutants complement all B mutants, but combinations of different A mutants or combinations of different B mutants fail to complement. In recombination tests, the A mutants map in one segment of the phage chromosome and the B mutants map in a separate, but adjacent, segment, as noted in Fig. 8.2. The *rII* region, like the bacterial tryptophan synthetase region (Chapters 4 and 5), comprises two discrete functional genes or cistrons: cistron *A* and cistron *B*.

A few *rII* mutants fail to complement with any A or B mutant in the *trans* test. Recombination tests show invariably that such aberrant mutants either suffer deletions that involve parts of both the *A* and *B* cistrons or are double mutants, having two separate single-site mutations, one in each cistron.

The nucleotides between genes

The one gene, one polypeptide chain theory proposes that each gene is a distinct entity. Yet, the gene string is a long, continuous polynucleotide chain composed of many genes. What then separates one gene from another?

In Chapter 7, operator regions were pictured as sites on the chromosome where mRNA synthesis is initiated. Special nucleotide sequences composing portions of individual operators also might serve as spacers between genes. These would be effective at the level of transcription

Fig. 8.2a. Evidence that mutant *r1589* is a true deletion of a spacer in the T4 *rII* region. (1) In recombination tests, two mutations, *a* and *b*, give a map distance, *n*. (2) The same two mutations are each combined with *r1589* (shaded portion). The double mutants are then crossed with each other. The map distance between *a* and *b* has decreased. It is now a distance of *n* (the original map distance) minus *d* (the extent of the deletion). Based on M. Nomura and S. Benzer, *J. Mol. Biol., 3* (1961), 684.

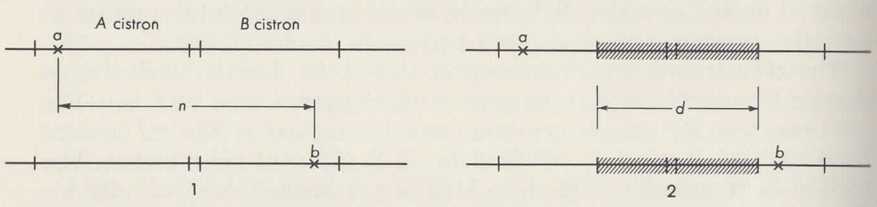
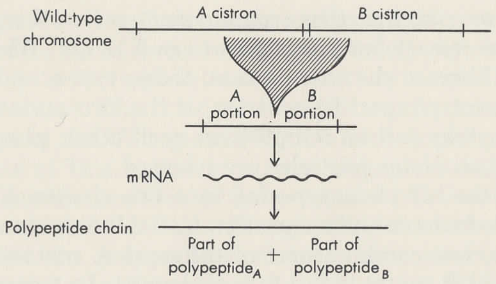


Fig. 8.2b. Interpretation of the effect of the *r1589* deletion of a spacer in the T4 *rII* region. Two originally distinct cistrons have become joined, resulting in a single gene product (polypeptide chain). Based on S. Benzer and S. P. Champe, *Proc. Nat. Acad. Sci. U.S., 48* (1962), 1114.



from DNA into the mRNA. This would be only a partial resolution of the problem, however, for a single operator region may control several adjacent, but autonomous, structural genes (Chapter 7). In the case of the histidine operon, the entire deoxyribopolynucleotide sequence within a complex operon appears to be transcribed into a single long mRNA molecule that remains unfragmented and yet is read discontinuously during protein synthesis.

What genetic data are available that imply the presence of spacers between genes? Some evidence comes from studies of the two cistrons, or genes, of the *rII* region of bacteriophage T4. The experiments rely upon the fortunate isolation and recognition of an unusual *rII* mutant, *r1589*. As shown in the genetic map (Fig. 8.2b), mutation *r1589* extends over portions of both the *A* and the *B* cistrons. That *r1589* is a deletion mutant has been demonstrated by the genetic test outlined in Fig. 8.2a. The map distance between mutations *a* and *b* is shorter when they are both combined with the *r1589* deletion. While most of the mutants whose genetic alterations involve both the *A* and the *B* cistrons lack both A and B function, *r1589* lacks only A activity. As shown by complementation in the *trans* configuration, it retains B activity. A most logical explanation of this behavior is that, although portions of both the *A* and *B* cistrons have been deleted, deletion of the terminal portion of the B cistron does not inactivate its product.

Further studies indicate that *r1589* lacks not only portions of the *A* and *B* cistrons but also a spacer originally present between these cistrons (Fig. 8.2b). Such an event has been postulated to lead to the construction of a continuous polypeptide chain of which part is B protein and part is A protein still specified by the residual *A* genetic region. The evidence for this suggestion is the unique location of those mutations that inactivate the remaining B protein activity in mutant *r1589*. In contrast to wild-type T4, where only those mutations located within the *B* cistron will inactivate B protein, some mutations in the *A* cistron of mutant *r1589* now lead to inactivation of the residual B activity.

The results of the *r1589* experiments can be interpreted most readily by postulating that a short, special nucleotide sequence acts as a spacer between the *A* and *B* cistrons of the T4 *rII* region. The spacer ensures the independent functioning of the two cistrons; without it, the two functions become interdependent or united.

Several situations in genetics are known that could be interpreted in a similar fashion. For example, among histidine-requiring mutants of *Salmonella*, one mutation appears to have joined two formerly independent enzyme activities. Mutant *22* (Fig. 4.5) contains a weakly active, giant protein in which transferase (gene *H*) and transaminase (gene *C*) activities may have become joined. Analysis of this and of other systems should assist us in comprehending how mRNA is translated into protein.

Making sense out of missense

Several (theoretical) examples of alterations in coding units are illustrated in Fig. 8.1. A careful study of the mutations and the possible coding triplets in the illustration will convince you that the primary mutant with the glutamic acid substitution can regain the wild-type phenotype either by backmutation to the original wild-type glycine coding unit or by mutation to a different missense coding unit, so that valine is inserted instead of glutamic acid. With this second mutation, the protein is also a mutant molecule, but because it is functional, the resultant gross phenotype may approach the normal condition. A functional protein also can be obtained by secondary mutation in a different coding unit in the same gene, and this phenomenon is known as second-site reversion or intragenic suppression (Fig. 4.4). A second missense mutation occurs in the same cistron and leads to the production of a protein with two amino acid substitutions. The second substitution sufficiently modifies the conformation of the protein to restore some enzyme activity. In fact, the above amino acid replacements and functional relationships have been found for *E. coli* tryptophan synthetase. (For review, see Chapter 4.) Here, the entire population of tryptophan synthetase molecules is of one kind; a particular amino acid always is located at a certain position in the polypeptide chain.

Now let us consider the somewhat different situation where certain secondary mutants contain a mixture of two structurally different tryptophan synthetase proteins: the original mutant molecules and, in addition, a small but appreciable proportion (from 1 to 10 per cent) of molecules that resemble the wild-type enzyme in structure and function. These secondary mutants grow slowly on minimal medium lacking tryptophan, but their growth is still markedly stimulated by the addition of tryptophan to the medium. The second mutation that partially suppresses the original mutant phenotype is located elsewhere in the chromosome, not in the tryptophan synthetase *A* or *B* genes. A secondary mutation of this type that can partially or completely neutralize the deleterious effect of a primary mutation is called an extragenic suppressor mutation.

How can suppressor mutations effect the production of small amounts of protein identical with or similar to the wild type even while the mutant protein continues to be formed? Present evidence suggests that the suppressor mutations of this type affect the translation of the genetic code in some manner that permits functional protein to be made in small quantity.

What are the possibilities of altering the translating mechanisms of protein synthesis? You might postulate that the strain carrying the suppressor mutation has a high rate of messenger mutations, but, if this were so, more than two types of proteins might be expected. You might theorize that the suppressor mutation occurs in the structural gene of an activating enzyme; such a mutation might alter the specificity of the activating enzyme, so that it occasionally activates the "wrong" amino acid (which, in the suppressed mutant, rectifies the original mutational injustice). Or you might theorize that the suppressor mutation is in a gene for sRNA and the altered sRNA has gained a low level of affinity for some activating enzyme that it normally does not react with. It is also possible that suppressor mutations act by altering some ribosomal structure that is required for accurate reading of the mRNA code.

Making sense out of nonsense

Nonsense coding units are not read, it is conjectured, because the cell lacks amino acid–sRNA molecules of the correct configuration to read these coding units during protein synthesis. For example, the nonsense triplet might be a triplet that is identical to or closely resembles one of the spacers between genes. Or perhaps the proper sRNA molecules are present but the existing activating enzymes lack the specificity for charging these sRNA's with their amino acids; or the concentration or affinity of the proper sRNA may be too low for effective translation. Conversely, the correct specific activating enzyme may be present, but the compatible sRNA absent. Consequently, mutation might initiate the formation of an activating enzyme with the specificity necessary to charge the unique sRNA, or it might modify the structure of the sRNA so that it reacts with the available activating enzyme. In these ways, mutations that affect the protein-synthesizing machinery could, hypothetically, convert nonsense into either missense or sense.

Such mutations have, indeed, been found, although indisputable biochemical evidence for the mechanism underlying the change again is still forthcoming. Certain *rII* bacteriophage mutants that ordinarily are unable to grow on *E. coli* K12 (λ) are characterized by their susceptibility to suppression by mutation of a specific gene in the host bacterium. Several characteristics are shared by this select group of mutants. Their phenotype is reversed by 5FU; and their genotype reverts to wild type after treatment with 2-aminopurine, but not with other mutagens. This class of *rII* mutants, which constitutes a few per cent of the total number of *rII* mutants, already has been discussed in reference to mRNA synthesis (Chapter 7). Furthermore, these *rII* mutants are included among those cistron *A* mutants that, when combined with the deletion mutant *r1589,* eliminate the activity of the *B* cistron (see earlier discussion in this chapter). For these reasons, this group of *rII* mutants is assumed to reflect a mutation that has effected the transcription of a particular nonsense configuration into an mRNA copy.

The same bacterial suppressor mutation that suppresses this group of *rII* mutants of T4 bacteriophage also suppresses some mutations in bacterial genes. Some of the suppressed mutants make nearly normal amounts of enzyme. In contrast to the missense mutants, the nonsense mutants produce no altered enzyme (CRM; see Chapter 4), and no CRM is produced in the suppressed strains along with the normal enzyme. Consequently, the most reasonable conclusion appears to be that the suppressor mutation has induced some corrective measure in one of the steps in the assembly of the protein so that the nonsense coding unit can be read. The reading may be one of missense, rather than of sense, because, although the suppressor restores full enzyme activity in some mutants, it restores only 3 per cent of normal activity in others.

Modulation

Recall the mention earlier in this chapter of the spacers between genes: short nucleotide sections that might allow the informational content of a giant messenger molecule to be read in sections, each section dictating the structure of an individual polypeptide chain. But even more information must be present, for the polypeptide chains dictated by the messenger are not present in the cells in equal amounts. Some sequences of the messenger appear to be read more often than others. The nucleotide sequences at the beginning of the molecule are translated more frequently than are some succeeding ones. This results in a greater number of polypeptide chains being formed for genes proximal than for those distal to the operator end of the operon. An example will help clarify this point.

Recall that histidine biosynthesis in *Salmonella* is controlled by nine genes, arranged in the order *E, I, F, A, H, B, C, D, G* (Fig. 4.5) and composing an operon. These genes function coordinately through the production of a messenger molecule representing the entire nine genes. It appears that the quantity of protein produced by gene *G* is slightly greater than the quantity of protein produced by gene *D*; and the latter is much greater than the quantity of protein produced, for example, by gene *F* or by gene *A*. Most important, the *ratios* of the amounts of the various proteins are invariant under conditions of repression and derepression. This tells us that, after the code is transcribed into a multigenic mRNA molecule, there is a certain constant probability that the entire messenger will be read, a higher probability that only the first half of it will be read, and an even higher probability that only the initial quarter of it will be read, and so on (modulation).

Support for the idea of modulation comes from further genetic studies. Most mutations in the histidine region enhance modulation. These mutations are detectable, not only because they eliminate the production of one active enzyme, but because they also decrease the amount of enzyme produced by each of the genes to the "left" of the mutant site. That is, the code in the part of the messenger representing genes nearer to the operator end of the molecule is read with normal frequency, but that part of the messenger representing the code of the distal genes is read less often.

Modulation of the translation process allows different amounts of coordinately produced enzymes to be formed from a single template molecule. Selection for appropriate modulating coding units and arrangements of genes within operons could result in the most effective enzymic balance for the survival of the cell. The existence of modulation strongly implies that the ribosomes read the multigenic messenger in a *unidirectional* fashion, attaching at one end (the operator end) and progressing along the mRNA chain. A number of ribosomes may read the same mRNA simultaneously, leading to the formation of polysomes, but each ribosome must start at the "beginning" end. The *arrangement* of genes in an operon is important, for the reading frequency is a polarized phenomenon.

What mechanisms underlie modulation? We can theorize that modulation is achieved in a rather nonspecific fashion; for example, modulation may result from the formation of some weak secondary structure in the mRNA molecule: any propensity for intrastrand base pairing would have to be overcome, the molecule would have to be stretched out, before it could be read. Alternatively, modulation might be achieved by the content of special coding units, modulating triplets, in the mRNA. Modulating triplets might be those for which charged sRNA is limited or cannot be positioned on the template readily. For example, we can postulate that the presence of a particular sRNA at the C-terminal end of the growing polypeptide chain affects the rate of placement of certain charged sRNA's on the succeeding triplet of the messenger.

Both of the above postulated mechanisms stipulate that modulation is achieved by slowing up messenger reading. This slowing-up could result in disengagement of the ribosome from the messenger or in destruction of the terminal portion of the mRNA beyond the point where translation was momentarily halted. Perhaps mRNA must be read in order to be preserved. Following this line of reasoning, we can postulate that an O^0 mutation (Chapter 7) is one that hinders reading and thereby allows destruction of the entire messenger.

Certainly, much work remains to be done in order for us to formulate a concrete picture of the translation process and its importance in regulation of gene function. We can expect many exciting advances in this area of genetic investigation during the next few years.

References

Benzer, S., "On the Topography of the Genetic Fine Structure," *Proc. Nat. Acad. Sci. U.S.*, *47* (1961), 403. This and papers preceding it have influenced our concepts of genetic fine structure and supply a system whose potential in facilitating new genetic discovery remains undiminished.

Brody, S., and C. Yanofsky, "Suppressor Gene Alteration of Protein Primary Structure," *Proc. Nat. Acad. Sci. U.S.*, *50* (1964), 9. A recent step in the unraveling of presumed genetic modifications of the protein-biosynthesizing machinery; note the combination of experimental approaches utilized.

Davies, J., W. Gilbert, and L. Gorini, "Streptomycin, Suppression, and the Code," *Proc. Nat. Acad. Sci. U.S.*, *51* (1964), 883. Suggests a role of ribosomal structure in the action of suppressor genes.

Dintzis, H. J., "Assembly of the Peptide Chains of Hemoglobin," *Proc. Nat. Acad. Sci. U.S.*, *47* (1961), 247. A cleverly designed experiment and analysis demonstrating the linear growth of the polypeptide chain during protein biosynthesis. Further work on this system is contained in *Informational Macromolecules,* cited at the end of Chapter 9.

Questions

8.1. What advantages does the *rII* system offer for genetical analysis? What disadvantages?

8.2. By what tests can one pursue an analysis to determine if a partial back-mutant is due to a suppressor mutation in a second gene, at a second site in the same gene, or at the same nucleotide pair involved in the first mutation?

8.3. What effect does a missense mutation have on synthesis of a protein? A nonsense mutation?

8.4. Tell what modulation is and describe two methods by which it is detected.

8.5. Outline a sequence of genetic events that you think might lead to the evolution of a new gene.

8.6. Read an original paper written during the past year on the mechanism of protein biosynthesis and correlate its salient points with the genetically oriented description in this and earlier chapters.

Nine

The Genetic Code

This chapter will bring together experimental evidence that outlines many general, and some specific, features of the genetic code.

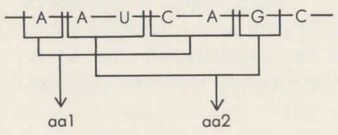
The generalization will be made that primary gene action consists of the transcription of the coded genetic information in DNA into the ribopolynucleotide sequence of mRNA and the subsequent translation of this information, contained in the four-letter alphabet, A, U, G, and C, of mRNA, into specific amino acid sequences. We will find that the four-letter alphabet is arranged into adjacent, nonoverlapping, three-letter words or coding units. Each coding unit in the mRNA molecule dictates the positioning of a specific amino acid in a polypeptide chain.

General types of codes

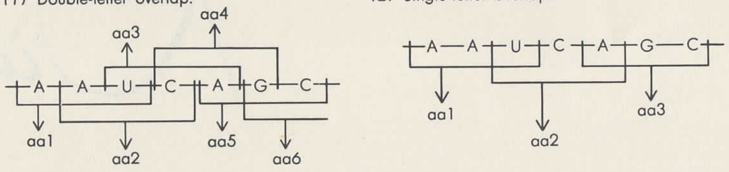
Figure 9.1 shows some theoretical types of codes. In 9.1a, discontinuous groups of nucleotides code for two amino acids, aa1 and aa2, that are adjacent to each other in the polypeptide chain. In 9.1b-d, the coding units are short, continuous sections of the polynucleotide chain; but in 9.1b the coding units overlap, while in 9.1c and 9.1d they do not. Overlapping means that a particular nucleotide serves in more than one coding unit. Finally, 9.1c and 9.1d differ in that 9.1c contains spacers (commas) between the coding units.

Fig. 9.1. General theoretical coding possibilities of mRNA containing the four bases A, U, G, and C.

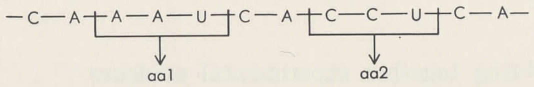
(a) Discontinuous.



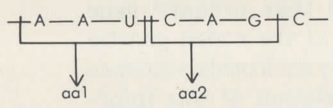
(b) Continuous, overlapping:
(1) Double-letter overlap. *(2)* Single-letter overlap.



(c) Continuous, nonoverlapping, with commas.



(d) Continuous, nonoverlapping, commaless.



The information presented in this chapter will rule out general models shown in 9.1a-c, while remaining compatible with the model shown in 9.1d. The coding units appear to be short, continuous, adjacent sections of the polynucleotide chain.

Deciphering the code

Three main types of experimental approaches have yielded most of the current information on the general nature of the genetic code and on the nucleotide composition of some of the coding units themselves.

The first method is analysis of the primary structures of wild-type and mutant proteins (Chapter 4). In each of the mutant proteins so far analyzed, a mutant amino acid replaces a particular wild-type amino acid. The accumulation of data on such amino acid replacement sets restrictions on the genetic code, as will be indicated in later sections of this chapter.

The second method is the use of synthetic ribopolynucleotides to stimulate amino acid incorporation in cell-free systems (Chapter 3). The polymers that effect incorporation of particular amino acids into polypeptide chains are considered to contain the code for those amino acids. Further refinements in the experimental procedure, outlined later in this chapter, seem capable of yielding information that can be used to establish the sequence of nucleotides in particular coding units. Some short-chain ribopolynucleotides can be synthesized by chemical means, but longer-chain synthetic ribopolynucleotides are made with an enzyme, polynucleotide phosphorylase, extracted from microbial cells. This enzyme differs from the RNA polymerases described earlier (Chapter 3) in several ways. It uses nucleotide diphosphates as substrates instead of the triphosphates used by the other polymerases. The equilibrium of the reaction is strongly in the direction of ribopolynucleotide breakdown rather than synthesis, although enough synthesis occurs to use the reaction for the preparation of polynucleotides in vitro. Most important, the composition of the polymer depends heavily upon the relative concentrations of the free nucleotides introduced into the incubation mixture rather than on the composition of a primer molecule. Consequently, while polynucleotide phosphorylase is inappropriate as a participant in a replicative or copying scheme and its function in vivo appears to be concerned with the breakdown of RNA, the enzyme is useful in allowing in vitro synthesis of various types of polymers whose composition is largely controlled by the nucleotide diphosphates placed in the reaction mixture.

The final means of delineating the genetic code involves inference from the results of purely genetical experiments. The experiments are of a variety of types. Some consider the activities of certain mutagens in reversing particular kinds of mutations; others involve placing two or more mutations on the same DNA strand and observing the resultant phenotype. Each kind of genetical experiment is described in the section most pertinent to the information it has provided.

The coding ratio

Recall that there are only four letters available in mRNA: A, U, G, and C. If the coding units for amino acids were composed of just two letters, there would be only 4^2, or 16, possible sequences to code twenty amino acids. Furthermore, there is strong evidence that more than one kind of coding unit can code a single amino acid (degeneracy). The shortage of coding units in a two-letter code would be circumvented were some other material, for example, the ribosomal RNA, to assist in the coding. But, as has been shown (Chapters 3 and 6), such assistance has been ruled out. Three-letter coding units offer 4^3, or 64, possibilities. Coding units larger than three letters would seem wasteful, but might still be worthy of consideration, except that accumulated evidence already has suggested that such larger units are unlikely.

The coding ratio is equal to the number of nucleotides in the mRNA (or nucleotide pairs in double-stranded DNA) divided by the number of amino acids in the polypeptide chain product; thus, it expresses the average number of nucleotides per coding unit. In each of several genes, the number of nucleotide pairs has been estimated and compared with the number of amino acid residues of the protein coded by the gene. All the estimates give coding ratios close to three; that is, three nucleotides compose a coding unit. Such a code is called a triplet code. A wide variety of genetic experiments is consistent with a triplet code, but measurement techniques are not precise enough generally to distinguish a ratio of exactly three from a ratio close to three (for example, two, four, or even five or six), except in the case of bacteriophage T4, where the estimate of three appears almost certainly to be valid. Additional types of evidence, to be discussed, suggest that the coding ratio is uniformly three for most, if not for all, amino acids.

Adjacent nucleotides

A simple genetic experiment demonstrates that the individual nucleotides of each coding unit are adjacent or very nearly adjacent to each other on the polynucleotide strand.

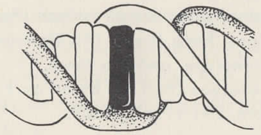
In a discussion of Fig. 4.4 it was pointed out that several tryptophan synthetase A protein mutants have been obtained that have different amino acid replacements at the same residue in the polypeptide chain. One mutant, *A23*, contains an arginine in place of the wild-type glycine, and another, *A46*, contains glutamic acid in place of the same glycine residue. The sites of the two mutations, *A23* and *A46*, are very close to each other but, nevertheless, are separable by recombination tests. Further correspondences between locations of mutations on the genetic map and locations of amino acid substitutions in mutant tryptophan synthetase proteins are shown in Fig. 1.3. These elegant studies demonstrate that the coding unit for one amino acid is composed of more than one nucleotide pair and that its member nucleotides are probably contiguous.

Triplet code

Some critical information on the nature of the coding unit evolved from studies of the mutagenic effects of acridines. A brief digression on the mode of action of an acridine, proflavin, is warranted here, for interpretation of the genetical data depends upon some understanding of the interaction of proflavin with polynucleotide chains.

Proflavin is a powerful mutagen when it is present during DNA replication. Figure 9.2 shows a model, based on physicochemical data, of the interaction of proflavin with native double-stranded DNA. The flat, planar proflavin molecules become intercalated between the stacked nucleotides of the double-stranded DNA, thereby stretching and stiffening the backbone structure of the DNA. Thus, proflavin is believed to exert its mutagenic effect by distorting the DNA molecule during replication or during recombination. Intercalation of a proflavin molecule between two nucleotides of one strand of the DNA almost exactly doubles the distance between them. This might allow the insertion of an extra nucleotide in the strand during replication or might shift the paired nucleotides by a single base during the events leading to recombination (that is, lead to unequal crossing over). Through one or both of these mechanisms, proflavin is believed to lead to the deletion or duplication of one nucleotide pair or several adjacent pairs. What genetical evidence supports such a model of proflavin-induced mutation?

Fig. 9.2. Acridines (black disk in the figure) are believed to intercalate between the base pairs of the DNA double helix, stretching the backbone and causing mutations through the deletion or addition of one or more base pairs. Based on L. S. Lerman, *Proc. Nat. Acad. Sci. U.S.*, *49* (1963), 98.



Proflavin-induced mutants revert to the wild-type phenotype spontaneously, and they can be induced to revert at high frequency with additional proflavin treatment. They do not revert to the wild-type phenotype after treatment with mutagens that cause replacements, rather than deletions or additions, of bases in the DNA molecule. (See the discussion of base-pair substitutions presented in conjunction with Fig. 7.3a.) Conversely, proflavin does not cause analogue-induced mutants to revert. The basic assumption, then, is that proflavin causes a base pair or base pairs to be deleted or duplicated.

In experiments with bacteriophage T4, spontaneous revertants were isolated from a proflavin-induced mutant of the *rIIB* gene. The reversions were not true backmutants to the original nucleotide sequence present in the wild-type DNA. By recombination tests with wild type, these revertants were proved to have a second mutation in the same short section (about a fifth) of the *B* cistron containing the original mutation. Individually, each of the two mutations elicits a mutant phenotype, but together they produce a wild-type phenotype. Because the secondary mutant sites are separable by recombination from the original site of mutation, but are located in the same gene, they are termed intragenic suppressor mutations. These are arbitrarily designated as "+" mutations in Fig. 9.3a, while the original mutation, *a*, is arbitrarily termed a "−" mutation.

Several of these secondary ("+") mutations were separated from the original mutation by recombination with wild type. Further reversions to the wild-type phenotype were isolated from the secondary ("+") single-site mutants. Once more, these revertants did not represent true backmutations at the original site but proved instead to be double mutants. The individual single-site mutants from this third class of double mutants also were obtained by recombination with wild type (bottom line of Fig. 9.3b).

Fig. 9.3a. Changes in the reading frame. The plus and minus signs denote independently isolated mutants of two classes; the two classes are isolated by selection for their ability to suppress the mutant phenotype of a particular mutant, arbitrarily labeled "—" (in which case the suppressor is denoted "+") or to suppress one of the "+" mutations (in which case the designation is "—"). This wild-type phenotype is indicated in Part IV of the table. Each of the mutations individually determines a mutant phenotype (Part I), as do any two "+" or any two, four, or five "—" combinations (Parts II and III). However, either three or six "—" mutations give rise to the wild-type phenotype, and any three "+" mutations, together on the same chromosome, give a wild-type phenotype (Part V). Based on F. H. C. Crick, L. Barnett, S. Brenner, and R. J. Watts-Tobin, *Nature,* *191* (1961), 1227.

Part	Genotype	Phenotype	Description
I	— or +	Mutant	Individual single-site mutations
II	— — or + +	Mutant	Double mutants of the same kind
III	— — — — or — — — — —	Mutant	Quadruple and pentuple mutants of the same kind
IV	— +	Wild type	Double mutants of the opposite kind
V	— — — or — — — — — — or + + +	Wild type	Triple mutants of the same kind or multiple of triple mutants of the same kind

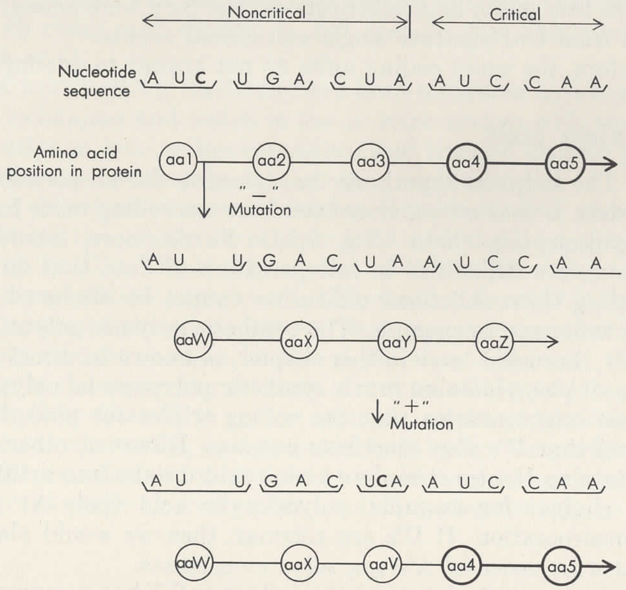
Hence, three classes of mutants were recognized: the original single-site *rIIB* mutant ("—"), suppressors of this original mutant ("+"), and suppressors of the suppressors ("—").

By genetic crosses, recombinants can be obtained having one, two, three, or more of the various "—" or "+" mutations in the same gene (*cis* configuration). The various combinations and their phenotypes are shown in Fig. 9.3. Two "—" mutations or two "+" mutations are still mutant, whereas most pairs of "—" and "+" mutations can form the wild-type phenotype. Furthermore, almost any three "—" mutations, or any three "+" mutations also form the wild-type phenotype. Six "—" mutations also form the wild-type phenotype.

The interpretation of these results is that the mutation shifts the reading frame of a triplet code (see Fig. 9.3b). If the code is thrown off by deletion of one letter, the subsequent triplets mainly form missense. This means that the triplet still codes for an amino acid, but it is the wrong amino acid for that particular position in the polypeptide chain. If the triplet reading frame is shifted one letter to the right, all subsequent triplets are read incorrectly until a second (suppressor) mutation occurs that can offset the first error by shifting the reading frame back one letter to the correct position. Therefore, as long as this second mutation occurs in the same cistron and near the original mutation, the small jumbled segment of the code located between the two mutant sites may cause damage insufficient to prevent formation of a functional gene product.

The suppressor mutations occur in a restricted portion of the *rIIB* gene. It is assumed that the amino acid sequence dictated by this portion of the *B* gene is not essential for *B* activity. (Recall from Chapter 8 that deletion of this portion of the genome in phage T4 can occur without inactivating *B* gene function.) Thus, it is important only that the reading frame be thrown back into correct position before the code for some critical portion of the amino acid sequence is copied.

Fig. 9.3b. Reading frame interpretation. In this diagram, it is assumed that the nucleotide sequence is read from left to right during protein synthesis and that the amino acids at positions 1, 2, and 3 are not critical to the function of the protein but that amino acids beyond position 3 are critical. The "—" and "+" mutations are assumed to move the reading frame so that coding triplets are missense or nonsense. The reading frame can be readjusted on center either by a second mutation compensating for the first mutation, as shown here, or by two additional mutations of the same kind as the original single-site mutation. Based on F. H. C. Crick, L. Barnett, S. Brenner, and R. J. Watts-Tobin, *Nature,* *191* (1961), 1227.



Let us emphasize several important conclusions from our discussion. First, the experiments demonstrate the effectiveness of groups of three similar mutations. This is strong evidence that the coding unit is a triplet. Second, the three nucleotide members of the triplet are closely adjacent in the polynucleotide strand and are very likely contiguous. The results also indicate, as will be discussed below, that the code is commaless, nonoverlapping, and highly degenerate, and is read in only one direction.

Nonoverlapping code

The sequence of amino acids in polypeptide chains appears to be almost completely random. This could not be so if coding units overlapped. For example, if the unit overlapped in two letters, the amino acid would always have the same four nearest neighbors (Fig. 9.1b1). Even if more than one unit can code for a particular amino acid (as appears to be so—see "Degenerate Code," below), this constancy in nearest neighbors would be found with a measurable frequency, and randomness in amino acid sequence would not obtain.

Furthermore, one can also see in Fig. 9.1b1 that a single mutation would affect the position of two or more adjacent amino acids in the polypeptide chain; but only single amino acid changes have been found to arise from single-site mutations (see Figs. 4.4 and 4.7 for some examples). Of a group of thirty-five nitrous acid–induced mutants that showed changes in tobacco mosaic virus protein, twenty-nine had proteins with two amino acid substitutions, but they were nonadjacent and resulted from two separate single mutational events.

Therefore, the small coding units do not appear to overlap.

Commaless code

The simplest hypothesis that accounts for all pertinent experimental data is that no commas exist between coding units for an individual polypeptide chain (Fig. 9.1d). Furthermore, introduction of commas causes difficulties in interpretation of data that do not arise by omitting them. All these difficulties cannot be discussed here; one of them will serve as example. The synthetic polymer polyuridylic acid (poly-U), discussed later in this chapter, is known to direct the incorporation of phenylalanine into a synthetic polypeptide, polyphenylalanine. One might assume that the coding triplet for phenylalanine is UUU and that U's also constitute commas. However, other polymers, not containing U, also stimulate amino acid uptake into artificial polypeptide chains: for example, polyadenylic acid (poly-A) stimulates lysine incorporation. If U's are commas, then we would also have to say that a sequence of A's can serve as commas.

Thus, our current picture of the code in mRNA is a sequence of nucleotides arranged in a series of triplets, each triplet forming the coding unit for a single amino acid. Although spacer nucleotides, or commas, do not occur within the code for an individual polypeptide chain, spacers, or periods, probably are present between the codes for different polypeptide chains made along different parts of the molecule.

Degenerate code

A convincing body of evidence indicates that the genetic code is degenerate; that is, several different triplets code for one amino acid. Some amino acids appear to respond to as many as four coding units.

The first suggestion that the code might be degenerate was obtained from microbial complementation studies and immunological analyses of mutant proteins. The *majority* of spontaneous mutations and an even higher proportion of mutations induced by agents that specifically elicit base-pair substitutions in the DNA give rise to mutants that possess cross-reacting proteins (CRM). The reasoning used is that most mutations that elicit the production of CRM involve only minor amino acid changes in the proteins, such as single amino acid substitutions. That is, most mutations seem to be missense mutations. To account for the high proportion of missense mutations, one must assume that most of the sixty-four possible triplets are effective in coding for amino acids. Since there are only twenty amino acids, some must be coded by more than one triplet.

Amino acid incorporation data also suggest that some amino acids are coded by several different triplets. For example, in the presence of phenylalanine, poly-U does not lead to appreciable leucine incorporation, yet the mixed polymers poly-UA and poly-UG each induce leucine incorporation. This suggests that several different triplets code for leucine; all codes may contain U, but specific ones also alternatively contain A or G.

Several leucyl-sRNA fractions have been separated by chemical and physical techniques and tested in the in vitro system with those polymers effective in leucine incorporation. One fraction showed preferential incorporation with poly-UA, another with poly-UG.

Because mixed polymers with higher U content elicit more leucine incorporation, two of the three code letters in the leucine triplet are assumed to be U's. The third member of the triplet can be either A or G. This and similar experiments indicate that the code is degenerate mainly for one of the three letters: as a consequence of this degeneracy, only *certain* changes in one of the three letters of the triplet may cause a change in the primary structure of the protein; changes from A to G or from G to A, for example, might go unnoticed. Such degeneracy in the coding system could well buffer the proteins of the cell against the effects of mutation! It would seem nature's own inimitable way of utilizing everything, including degeneracy, to her advantage!

The degeneracy may not be total

There are two reasons for suggesting that not all of the sixty-four possible sequences actually are used for coding amino acids in a single organism. The first is our basic desire for uniformity. Recall from previous discussions the need for preserving some nucleotide sequences as critical portions of operators (Chapter 7) and as spacers between codes for different polypeptide chains (Chapter 8). The other reason stems from experiments with *rII* mutants of bacteriophage T4. In discussing reading-frame mutations earlier in this chapter, it was stated that *almost* all pairs of "−" and "+" mutations gave rise to the wild-type phenotype, but the qualification was not explained. Now we can say that one explanation for the exceptions could be the creation of nonsense triplets at specific points when the coding frame is shifted. A nonsense triplet could be one that is not read by sRNA (for example, Fig. 8-1). The combinations of "−" and "+" mutations that do not restore wild-type phenotype in the *rIIB* cistron constitute but a minority of the "−" and "+" pairs tested; the distribution of the mutations in question suggests that the code is highly degenerate but that rare coding triplets might not be read.

A further argument for the existence of nonsense triplets derives from experiments using the T4 *rII* mutant *r1589,* which lacks the spacer between the *A* and *B* cistrons (Chapter 8). In this mutant, where the cistrons appear to have become joined into a single functional unit, several types of mutations in the *A* region affect the B activity. Certain of these mutations, including proflavin-induced mutations, are deletions of various sizes. They are assumed to alter the reading frame throughout the functional unit. Other mutations are interpreted as being substitutions that create nonsense triplets. They are revertible by 2-aminopurine, which, as was mentioned before, causes base-pair substitutions (see Fig. 7.3).

Quasi-universal code

Can we integrate evidence for the code from different organisms? Is the sequence that codes for a particular amino acid in a bacteriophage protein precisely the same sequence that codes for the same amino acid in, let us say, a protein of the human? The answer to these two questions is a tentative yes, with, as we shall see, one qualification.

The best evidence for a general concept of universality in coding is that the products formed by in vitro ribosomal preparations obtained from different organisms always reflect the source of the mRNA used. In Chapter 3, you will recall, we mentioned that when amino acid–charged sRNA from the bacterium *E. coli* was placed with rabbit reticulocyte mRNA and ribosomes, the product was rabbit hemoglobin. By tryptic peptide analysis (Chapter 2), the hemoglobin was shown to be identical in its primary structure with hemoglobin made with reticulocyte sRNA. The bacterial sRNA appears to be able to recognize perfectly the mammalian mRNA for hemoglobin and also to work harmoniously with mammalian ribosomes.

Studies with synthetic messengers give credibility to a universal code. Bacterial ribosomes and sRNA's will participate in polyphenylalanine synthesis if poly-U is used; rabbit reticulocyte ribosomes and sRNA's will also make polyphenylalanine if poly-U is used.

Studies in vivo also support the contention of a universal code. Widely different strains of bacteria may exchange genetic material. The genes transferred into a new, distant species retain their autonomy. The protein product controlled by these genes is synthesized in a foreign environment, yet it appears to be identical to the protein formed in the homologous bacterium.

Exceptions to universality in genetic coding occur in certain bacterial mutants. In Chapter 8, mutants were described that appear to have gained the ability to read nonsense mutations. Such bacteria appear to differ from other organisms. Because of this, the code cannot be considered as completely universal—there may be rare exceptions.

The compositions of triplets: Amino acid incorporation studies

The stimulation of amino acid incorporation by synthetic polymers has allowed estimates of the base compositions of many coding triplets. By altering the ratios of the bases within a given type of polymer (for example, poly-UC), one can obtain an estimate of the composition of the triplet (for example, two U's and one C, as opposed to one U and two C's). The decision as to which of the two triplets is assigned is derived from observations on the quantitative efficiencies of a series of polymers in stimulating amino acid incorporation (Fig. 9.4). The stimulation of incorporation of each amino acid by different polymers is usually compared in relation, for example, to their efficiencies in stimulating the incorporation of phenylalanine (which is coded by UUU triplets).

The polymers serve another useful purpose; they limit selection of coding letters. A polymer that codes for several amino acids but not for others may be considered as lacking the coding units for the latter amino acids.

There are many pitfalls in methodology with the above synthetic systems. Incorporation of nucleotides into synthetic ribopolynucleotide polymers is not always random, and one must directly ascertain the base content and the average distribution of bases in the polymers in order to have a reliable estimate of the frequency of occurrence of particular triplets. Secondly, aberrant incorporation may occur. For example, the stimulation both of phenylalanine and of leucine incorporation by poly-U embarrassingly indicated that both amino acids might be coded by the same triplet. However, because leucine incorporation is appreciable only in the absence of phenylalanine and when abnormally high concentrations of magnesium ion are present, it is assumed that UUU codes for phenylalanine and that the leucine incorporation is an artifact. Other ambiguities may arise from impurity of radioactive amino acids, metabolism of the amino acids in the impure

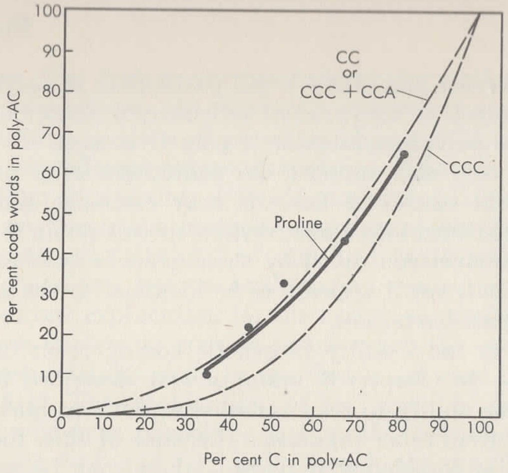


Fig. 9.4. Determination of the composition of coding triplets through amino acid incorporation efficiencies with various polymers. The actvities of a series of poly-AC preparations with different proportions of bases were tested in an amino acid–incorporating system. The theoretical proportions of the four doublet (CC, AC, etc.) and eight triplet (CCC, CCA, CAC, etc.) permutations were calculated for each of the polymers. The broken lines in the figure show theoretical frequencies of some of the RNA code words. The observed relative abilities in amino acid incorporation are shown, for proline, by the solid line. It is concluded that the doublet CC, or the two triplets CCC *and* CCA code for proline.

The four doublet permutations do not contain enough specific information to code for the six amino acids incorporated in experiments similar to that shown here with the poly-AC templates, whereas the informational content of the eight triplet words is adequate. The per cent incorporation of lysine, asparagine, glutamine, and histidine agrees well with triplet code-word frequencies, but not with doublet frequencies. Redrawn from M. W. Nirenberg, O. W. Jones, P. Leder, B. F. C. Clark, and S. Pestka, *Cold Spring Harbor Symp. Quant. Biol.*, 28 (1963), 549.

Fig. 9.5. Summary of mRNA coding triplets derived from amino acid incorporation data. * indicates probable, but tentative, code word. The nucleotide sequence of the triplets is arbitrary. From M. W. Nirenberg, O. W. Jones, P. Leder, B. F. C. Clark, and S. Pestka, *Cold Spring Harbor Symp. Quant. Biol.*, 28 (1963), 549.

Amino acid	Triplets
ala	CCG, CUG*, CAG*
arg	GCC, GAA, GCU*, GCA*
asp-N	AAC, AAU, ACU*
asp	GUA, GAC*, AAG*
cys	GUU
glu	AAG, AUG*, ACG*
glu-N	ACA, AGA, AGU*
gly	GUG, GAG, GCG
his	ACC, AUC*
ileu	UUA, UAA
leu	UGU, UCU, UAU, CCU
lys	AAA, AUA
met	UGA
phe	UUU, CUU
pro	CCC, CUC, CCA, CCG*
ser	UUC, UCC, UGC*, AGC*
thr	CAA, CAC
try	UGG
tyr	AUU
val	UUG, UAG*

extracts used for incorporation (for example, conversion of glutamine to glutamic acid), and alterations in the polymers by polynucleotide phosphorylase or other enzymes present in the extracts.

The amino acid–incorporating system allows assignment of code letters (Fig. 9.5) but does not reveal the order of the letters in the triplet. The order may be determined eventually by more precise use of the system. For instance, by strategic arrangement of the bases in short polymers and analysis of the polypeptide products, one may be able to define the sequence of bases in effective coding triplets. Recently, another approach has been developed in an attempt to analyze the sequence of code letters within coding triplets. It has been established that a trinucleotide of known sequence binds to ribosomes and specifically stimulates the binding of a particular sRNA charged with a radioactive amino acid. Through the use of a series of trinucleotides, each containing a different sequence, it may be possible to elucidate the detailed genetic code.

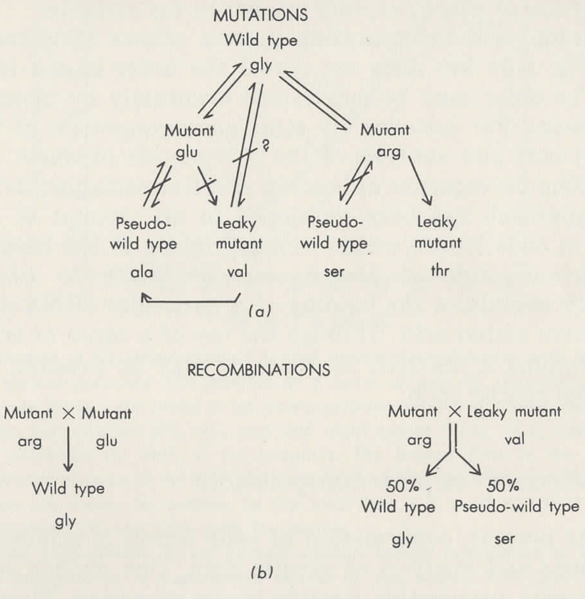
Verification of triplet composition

At present, confirmation of code letters is achieved by chemical methods and analysis of genetic data. One method is the use of mutagens with presumably specific modes of action. These have been used to induce tryptophan synthetase mutations, and the A proteins have been extracted from the mutants and screened for amino acid substitutions. As indicated in Fig. 9.6a, chemical mutagens that predominantly cause transition mutations (purine for purine and pyrimidine for pyrimidine substitutions) can effect certain amino acid substitutions (shown by open arrows) but cannot effect others (slashed arrows).

Genetic data from these same mutants have been analyzed. Some amino acid substitutions in the A protein are presented in Figs. 1.3 and 4.4 and summarized in Fig. 9.6a. At the position of a single residue, various A proteins contain either gly, glu, val, ala, arg, ser, or thr! The arrows in Fig. 9.6a show the various substitutions that have been demonstrated to take place by single mutational steps. However, each amino acid can be replaced by only a limited number of other amino acids. Members of the glu-val-ala group are interconvertible by mutations that are assumed to affect a single base pair in the DNA, and they do not mutate directly to members of the arg-ser-thr group. This is evidence that the mutations in the two groups affect different bases in the coding triplet. These mutational data on the A protein place further restrictions on the possible compositions of the coding triplets.

Recombination tests (Fig. 9.6b) support the conclusions drawn from the mutation studies. When mutants whose A protein contains arg are crossed with mutants whose protein contains glu, wild-type recombinants (whose protein contains gly) are obtained. This indicates that the code letter that has been changed (mutated) to give the arg substitution is different and recombinationally separable from the nucleotide

Fig. 9.6. Amino acid replacements at a single position in the primary structure of tryptophan synthetase A protein and the phenotypes of the strains.



pair that has mutated to give rise to the glu substitution. In contrast, the mutant containing glu fails to give wild-type recombinants when it is crossed with the mutant whose protein contains val. This indicates that the same base pair is differentially affected to elicit the glu and the val substitutions.

The cross between the arg mutant and the val mutant in Fig. 9.6b shows that the arg and val triplets contain recombinationally separable code letters, which, when assorted properly, code for gly and for ser. The recombinant product depends upon where in the triplet the recombinational event has occurred. Furthermore, because the two types of recombinants occur with low but approximately equal frequencies, they both are assumed to be caused by the same type of recombinational event. One would suspect that this is a single crossover.

Assignments of triplets

Figure 9.5 presents a current series of coding triplets that has been calculated from amino acid incorporation data. Because poly-U elicits phenylalanine incorporation, poly-A lysine incorporation, and poly-C proline incorporation, we can assign three triplets, UUU, AAA, and CCC, respectively. The problem of triplet assignment now increases in complexity. Assuming that the genetic code is universal, can we assign the triplets established by amino acid incorporation studies

(Fig. 9.5) in a manner that can account for genetic data (amino acid replacements in tobacco mosaic virus [Fig. 4.7] and tryptophan synthetase [Fig. 9.6]), as well as recombination and mutation data?

A large amount of data can be accounted for by making coding assignments, but there are discrepancies. Perhaps the code is not completely universal? Perhaps the code is not based on the standard rules of Watson–Crick base pairing as is currently assumed? More likely, we believe, some pieces of data have been incompletely or incorrectly interpreted and will not stand up to more critical testing in the future. Our current level of understanding of the intricacies of the amino acid–incorporating systems and of the modes of action of chemical mutagens indicates the likelihood of many further improvements and refinements. You, the patient reader, through perusal of the current literature, can follow the progress and excitement of deciphering the genetic code.

References

Bennett, J. C., and W. J. Dreyer, "Genetic Coding for Protein Structure," *Ann. Rev. Biochem.*, *33* (1964), 205. A recent review.

Chapeville, F., F. Lipmann, G. von Ehrenstein, B. Weisblum, W. J. Ray Jr., and S. Benzer, "On the Role of Soluble Ribonucleic Acid in Coding for Amino Acids," *Proc. Nat. Acad. Sci. U.S.*, *48* (1962), 1086.

Cold Spring Harbor Symp. Quant. Biol., *28*, "Synthesis and Structure of Macromolecules" (1963).

Crick, F. H. C., L. Barnett, S. Brenner, and J. R. Watts-Tobin, "General Nature of the Genetic Code for Proteins," *Nature*, *192* (1961), 1227. A classic paper in genetics is this presentation of genetic evidence for a triplet code.

Nirenberg, M. W., and J. H. Matthaei, "The Dependence of Cell-Free Protein Synthesis in *E. coli* upon Naturally Occurring or Synthetic Polyribonucleotides," *Proc. Nat. Acad. Sci. U.S.*, *47* (1961), 1588. The original paper on use of synthetic polyribonucleotides in amino acid–incorporating systems and their applications to analysis of the genetic code. Note that simple innovations in techniques in stabilizing the cell-free system greatly facilitated the experimental research.

Vogel, H. J., and V. Bryson, eds., *Informational Macromolecules*. New York: Academic Press, Inc., 1963.

Questions

9.1. One problem encountered in genetics is its specialized jargon. Furthermore, in a conceptual science, the same word may hold different meanings to different workers. The term *gene* has held different conceptual meanings to geneticists during the development of genetics this past half-century. At one time or another it has been used to describe a unit of mutation, a unit of function, and a unit within which recombination cannot occur. The following table shows the relationship between some terms of genetic jargon used in this text and their definitions. Fill in the chemical (DNA) equivalent description for each term. Find in the literature and in other genetics textbooks some synonyms of these terms, setting them out in tabular form as a continuation of the table below.

	Smallest unit of function		Smallest unit of mutation	Smallest unit indivisible by recombination
	In terms of mRNA	In terms of a polypeptide chain		
Chemical				
Genetical (this text)	*operon*	*gene*	*site*	*site*
Genetical (other texts)				

9.2. In a scientific journal (*Proc. Nat. Acad. Sci., U.S.; Science; J. Mol. Biol.;* etc.), search for a recent article on the genetic code. Write a one-page report in which you compare the observations in the article with the description in this chapter.

9.3. What is the evidence for a triplet code?

9.4. Look back at earlier chapters of this book and outline the reasons for believing that mRNA carries the genetic code.

9.5. List the reasons for believing that the genetic code is degenerate.

9.6. Some workers believe there is some form of life on the planet Mars.
(*a*) If you could visit Mars, what type (or types) of hereditary material, if any, might you expect to find? Why?
(*b*) Do you think the genetic code will be truly universal (or, at least, interplanetary)? Why?
(*c*) How would you go about testing to find out the chemical nature of the hereditary material in organisms from Mars?
(*d*) Assuming that the genetic material of a Martian organism is found to be DNA, how would you go about testing to see if the genetic code is (or is not) universal?

9.7. As outlined in this text, current theories of gene action require that there be spacers. Describe the positions at which such punctuation is necessary in the reading of the genetic code and speculate on the chemical nature of the punctuation in each case.

9.8. List in summary fashion the main methods that have been used in deciphering the genetic code.

Ten

Perspectives and Horizons

In the past, our ideas of the nature of the hereditary material and how it controls the characteristics of the organism have continually undergone revision. Certainly, our present concepts will similarly be tested and changed as more experimental information becomes available. The preceding chapters, then, have merely "frozen" within the confines of the printed page our comprehension of genic activities at the time of writing. Our documentation, while very incomplete, exemplifies the kinds of experimental evidence that have been obtained by biologists in support of the principles that underlie our current concepts. Some principles are thoroughly proven, but others require extensive further investigation before they can be accepted without reservation by either the reader or the authors.

In our discussions, we have emphasized the biochemical events at the intracellular level, for only here will a comprehension of the basic "whys" and "hows" be found. Once these general concepts are discerned, the reader will have no difficulty in applying his knowledge to interpretation of more classical genetical results. For example, Fig. 10.1 schematizes the genotypes and phenotypes of progeny from particular types of genetic crosses in diploid organisms. Some interpretations of the genetic data are suggested at the bottom of the figure. While this is a presumptive analysis, to be sure, it can provide a clue as to where to look for a mechanism in gene action at the biochemical level.

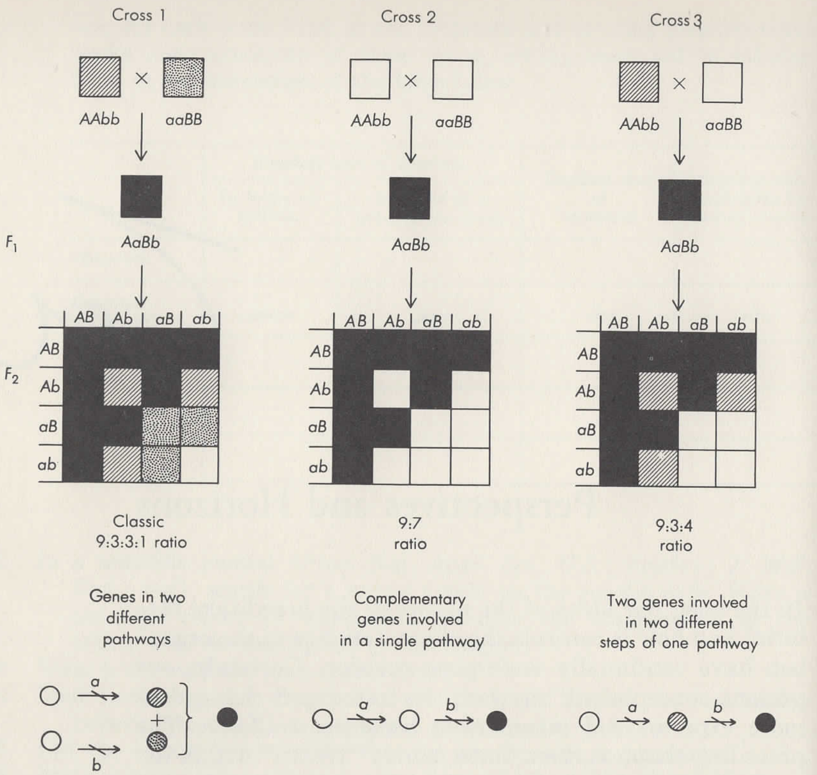

Cross 1
Cross 2
Cross 3
AAbb
aaBB
AAbb
aaBB
AAbb
aaBB
F1
AaBb
AaBb
AaBb
F2
AB
Ab
aB
ab
Classic 9:3:3:1 ratio
9:7 ratio
9:3:4 ratio
Genes in two different pathways
Complementary genes involved in a single pathway
Two genes involved in two different steps of one pathway
a
b

Fig. 10.1. Clues in genetic data that can assist further study of the mechanism of gene action. Three different crosses between diploid organisms heterozygous for two recessive genes (*a* and *b*) in the F_1 generation. The three cases lead to different Mendelian ratios (dihybrid ratios) in the F_2 generation; the genotypes and phenotypes are indicated for the F_2 generation in the Punnett squares. Biochemical interpretation of the data is indicated beneath each cross.

The classic 9:3:3:1 dihybrid ratio can arise from the relative independence of genetic blocks in biochemical reaction sequences that involve independent phenotypic traits. Sometimes the products of the reactions may be observed by their phenotypic effects in the same cells, tissue, or organ. The eye colors of insects and the pigmentation of portions of plants are cases in point, each containing a mixture of reaction products of relatively independent biochemical pathways. The other crosses shown in the figure demonstrate deviations from the classic dihybrid ratio that are, however, explicable by independent genes that control independent, enzymatically catalyzed reactions.

The eternal “why?”

The above brief examples suggest one means by which biochemical-genetic reasoning may assist in unravelling the secrets underlying the genetic control of phenotypes of higher organisms. Genetic concepts and model systems developed in microbial investigations often can aid the biologist whose major interest is higher organisms. The universality of many biological processes affords numerous opportunities for interdisciplinary “cross-feeding.”

It often has been noted that opportunities are afforded mainly to the curious. There are those who will not be satisfied simply with the knowledge that a 9:3:3:1 ratio is obtained in a particular situation and who will wish to delve more deeply into the problem of gene action by asking “Why?” When the “why” is answered at one level of analysis, the experimental biologist will often again ask “Why?” or “How?” and once more will be lured into new areas of exploration. This fundamental dissatisfaction with the unknown is a major driving force in all of science. Contemporary genetics certainly provides some outstanding examples of rewarding answers to well-chosen questions.

Growth and development

One of the major areas of biology now being infused with contemporary molecular genetic thought is the growth and development of higher organisms. How does the single fertilized egg become differentiated during subsequent cell divisions into the diverse specialized tissues that constitute the adult organism?

A vast amount of *descriptive* knowledge is available about the changes exhibited by organisms during development, and now, in addition, biochemical and genetical data are being applied to the problem (see Clement Markert’s *Developmental Genetics* in this series and the companion Developmental Biology Series).

A critical aspect of growth and development appears to involve the turning on and turning off of certain genes in particular groups of cells at very specific times. The techniques and approaches worked out for microorganisms, and described in the preceding chapters, are now being employed to determine whether this is so. For example, do certain cells contain a unique assortment of mRNA’s during a particular stage of development? Are there selective control mechanisms at other levels, such as in the rate of mRNA destruction, the availability of ribosomes, and the quantity, function, and specificity of certain types of sRNA? We know relatively little about cellular regulatory mechanisms in microorganisms; the major facets have been mentioned in earlier chapters of this book. We do not know whether these mechanisms will apply to higher organisms. And, once a parallel is found, one still must ask whether the mechanism is important in differentiation and how it operates.

In *The Mechanics of Inheritance,* Franklin W. Stahl mentions the structural differences between the chromosomes of bacteria and viruses and the chromosomes of higher organisms. The former are essentially naked DNA molecules, not surrounded by a nuclear membrane or associated with a variety of protein components, as are the latter. Do these proteins of chromosomes, histones and nonhistone proteins, play general or specific roles in gene action? What is the meaning and function of the nucleolus, an intranuclear structure rich in RNA and associated with a particular gene locus implicated in synthesis of ribosomes of higher organisms?

Intuition tells us that some similarities will be found between regulatory systems in microorganisms and those in higher organisms and that such systems will be found to be important in guiding differentiation. But science relies upon experimental demonstration, not on intuition. It is certainly possible that controlling mechanisms peculiar to higher organisms will be discovered. We do not yet have a detailed picture of the structure of chromosomes of higher organisms (see Stahl's *The Mechanics of Inheritance*); and without this, our understanding of how genic expression can be influenced and regulated is limited. For instance, the two genetical phenomena discussed below present characteristic enigmas. Neither phenomenon is understood, although many examples have been recorded.

Gene dosage compensation

In Chapter 5, Complementation, it was stated that gene products are present in proportion to gene dosage; the heterozygote produces one-half normal protein and one-half mutant protein. This is generally true of autosomal genes and their products, although exceptions to the rule constitute cases for conjecture. The rule often does not apply to sex chromosomes. The human female has two X chromosomes, while the male has only one X chromosome, along with a Y chromosome bearing relatively few genes (see McKusick's *Human Genetics*). Yet, where certain genes on the X chromosome are lacking on the Y chromosome, equal amounts of the proteins coded by those genes appear to be synthesized by males and by females. This is called dosage compensation.

In somatic cells of human females, one of the two X chromosomes seems to remain in a functional, extended state during interphase. The second X chromosome clumps, is afunctional, and forms a darkly staining chromatin body at the periphery of the nucleus (Fig. 10.2). Which X chromosome clumps is usually random, although, occasionally, heterozygous females are found in which only a particular one of the two X chromosomes functions in the vast majority of cells. Thus, the human female is normally a cellular mosaic, some cells possessing a functional X chromosome derived from the father, while others possess a functional X chromosome derived from the mother. The mechanism by which one of the two chromosomes becomes clumped (and, presumably, is turned off *in toto*) is unknown. Whatever the mechanism, only one X chromosome can function in a single cell (see McKusick's *Human Genetics*).

Dosage compensation is known also in *Drosophila*, although outright clumping of X chromosomes has not been described. Furthermore, in *Drosophila*, genetic evidence indicates that the function of the entire X chromosome is not turned off. This is illustrated by the lack of variegation of most sex-linked traits on the cellular level.

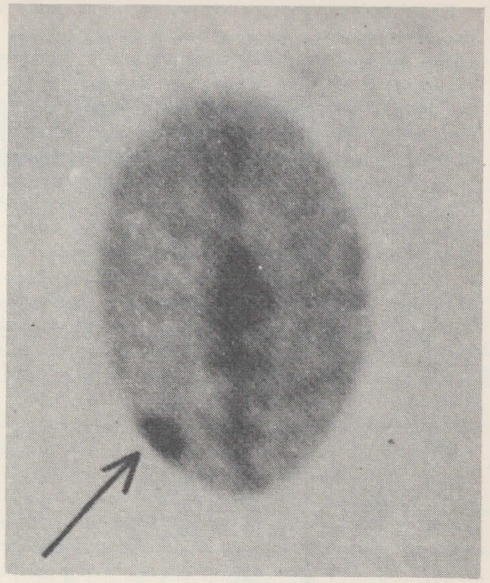


Fig. 10.2. Interphase nucleus from a cell of a normal human female, showing the peripheral body of chromatin that is presumed to represent a condensed X chromosome.

Variegated position effects

A gene adjacent to a region of the chromosome where breakage has occurred is sometimes stably mutant. This might be expected if the break has occurred within the polydeoxyribonucleotide segment of that gene, thus disrupting the base sequence requisite for normal function. This behavior is characteristic of chromosome breakage in higher organisms where the broken end is often rejoined to a lightly staining (euchromatic) portion of the same or of a different chromosome (*264-48, -53,* and *-63* in Fig. 10.3). Only occasionally is a discrepancy found. In *Drosophila* strain *264-34* (see Fig. 10.3), a stable mutation is produced even though the chromosome break has occurred some distance away from band *3C7* where the facet-Notch gene (*fa*) is supposedly located.

A contrasting behavior is exhibited when the end of a chromosome broken in a presumedly similar manner is rejoined to the darkly staining, basophilic sections of the chromosome (the heterochromatic regions, lower portion of Fig. 10.4). Now the genes over relatively long segments of the chromosome function only sporadically. The organism is a mosaic of phenotypically normal and mutant portions —it is variegated. Figure 10.4 is an illustration of this variegated position effect.

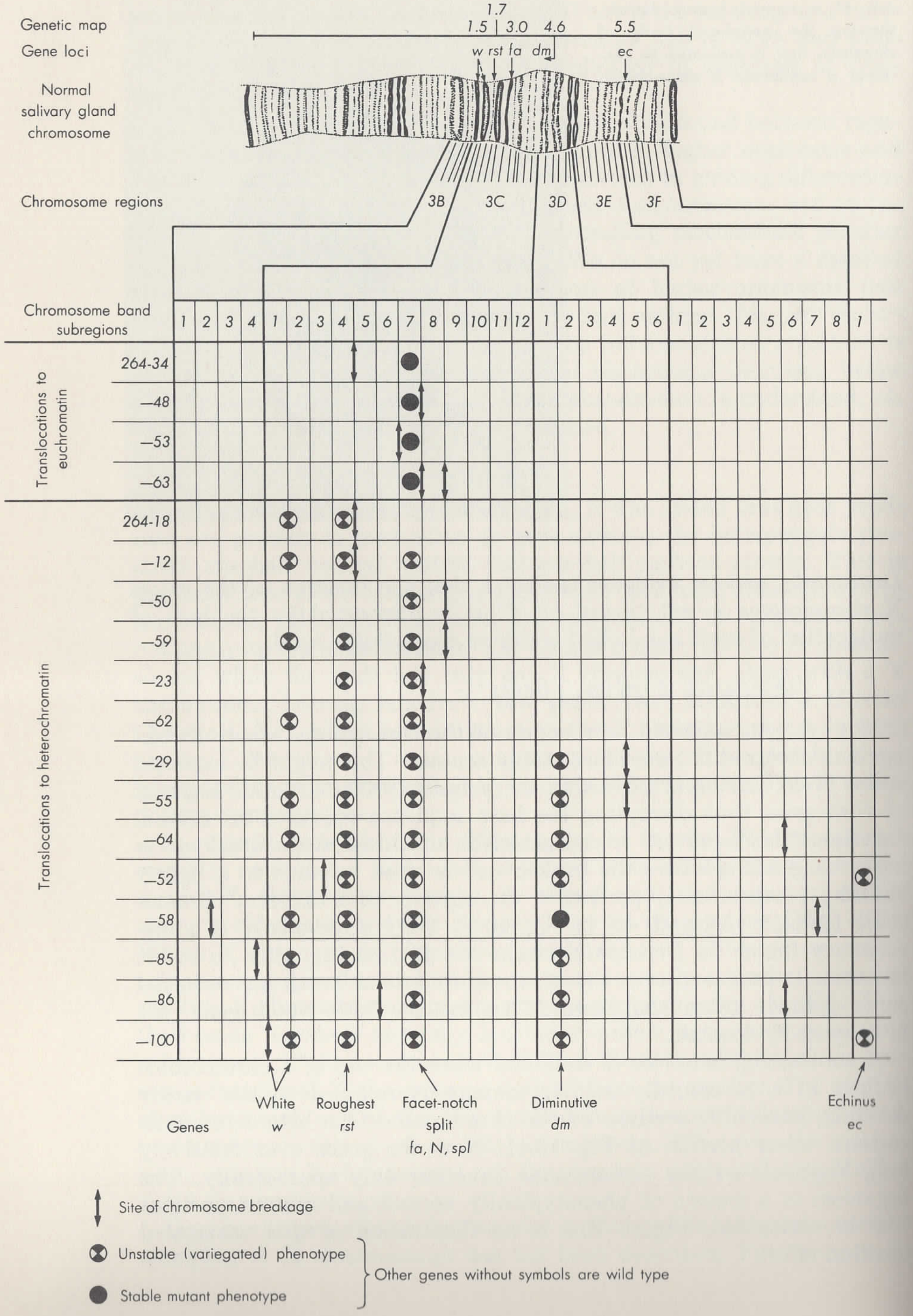


Fig. 10.3 (*opposite*). Variegated position effects in *Drosophila melanogaster* caused by aberrations involving the X chromosome. The positions at which five gene loci have been mapped are indicated at the top of the figure. The correlation of these gene loci with positions on the banded salivary gland chromosome is indicated beneath the genetic map (for methods, see Carl P. Swanson, Timothy Merz, and William J. Young's *Cytogenetics* in this series). Cytologists have agreed upon nomenclature, dividing the cytological structure into regions and subregions, a few of which (*3B1* through *3F8*) are indicated beneath the drawing of the normal salivary gland chromosome structure. A large number of translocations, inversions, and other chromosomal aberrations are known that involve this region of the X chromosome. The positions of the breaks that have given rise to these aberrations have been located by cytological analysis of the salivary gland chromosomes and are indicated by the vertical arrows. Other breaks, occurring outside the indicated region, are not shown in the figure.

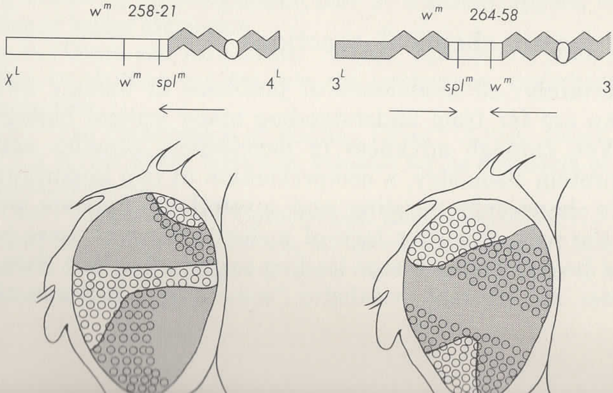
When the chromosome segment is translocated to a lightly staining, banded euchromatic portion of the X chromosome (or to a similarly staining region of another chromosome), the gene adjacent to the position of the break is sometimes stably mutant (● in the top section of the chart). Rarely, a gene several bands distant from the break is stably mutant (for example, in mutant *264-34*).

When a chromosome segment is translocated to a heterochromatic portion of the chromosome, genes quite distant from the position of the break cause a variegated phenotype by turning on and off during growth and development (⊗ in the lower section of the chart.) The number of genes affected depends upon the particular heterochromatic region to which the chromosome segment has been translocated. The gene nearest the heterochromatin (nearest the break) is turned off more frequently than are those further distant. Whenever a more distal gene is turned off, all of the intervening genes are also turned off.

When the giant salivary chromosomes of variegated flies are examined, the bands containing the genes undergoing variegation are sometimes clumped in the heterochromatic chromocenter rather than in their normal distended and banded form. Based in part on M. Demerec, *Genetics, 25* (1940), 618, and *Proc. Intern. Congr. Genet. 7th, 1941*, p. 100.

Fig. 10.4. The upper portion shows two translocations, *258-21* and *264-58*, that give rise to mottled eye color (w^m) in *Drosophila*. In *258-21*, part of the left arm of the X chromosome, shown in white (X^L), has been translocated to the fourth chromosome, and in *264-58* a smaller fragment has been translocated to the third chromosome. The positions of the chromosome breaks and genes affected in *264-58* are shown in Fig. 10.3.

The lower portion of the figure shows examples of the eyes of flies carrying one of the two translocations. The circles indicate the areas of rough-eye phenotype caused by the split gene; shaded areas indicate pigment, and white areas represent no pigment. Note that in translocation *258-21* the white areas are always rough. The position effect is exerted in one direction, from the position of the break (arrow above eye). In translocation *264-58* note that the white areas may or may not be rough and that the rough areas may or may not have pigment. The position effect is exerted independently from either side, because heterochromatic material is now present on both sides of the short translocated fragment. Redrawn from W. K. Baker, *Am. Zoologist, 3* (1963), 57.



The genes influenced by this position effect are not themselves structurally defective. They still are able to function in some cells of the organism, and they become stable wild-type genes once they are removed from the heterochromatic region, either by recombination or by a second chromosomal aberration. These observations indicate that the lack of activity of the genes is due to their position next to heterochromatin, and not to structural defects within the genes themselves. Some blocks of heterochromatin appear to exert strong position effects, while others exert weak effects. In addition, there is a clonal distribution of the phenotypes of cells in the mosaic individuals (Fig. 10.4); that is, the different types of cells occur in discrete groups.

Many speculations have been advanced to explain these observations; for the most part, they are inadequately supported by experimental evidence. Perhaps an inquisitive reader may someday furnish the answer to this or to other unanswered problems in the genetics of growth and development. What is the organization of the chromosome of higher organisms? What is heterochromatin? How does juxtaposition with heterochromatin elicit such striking instability in expression of the function of the gene?

Enzyme structure and function and the evolution of the protein molecule

In spite of considerable knowledge about the structure of proteins, we know relatively little about the molecular mechanisms and structural requirements underlying the catalytic activities of enzymes. Chemists have made some progress in structurally modifying proteins, but, undoubtedly, genetic modification—that is, the use of mutant proteins—offers a new and almost inexhaustible source of enzymes that are modified specifically at various positions in the primary structure. Such precise alteration in structure, without the attendant problems of denaturation and secondary modification characteristic of chemical or physical treatment, should provide a wealth of information, not only on the relationship between protein structure and function but also on the exact structural requirements for the maintenance and evolution of catalytic activities.

Vitalism versus chemical reactions

Certainly, an abundance of problems in biology await solution, for we are far from understanding many critical biological phenomena. Yet, through advances in biochemical genetics and nucleic acid and protein chemistry, a comprehension of two important properties of life, heretofore puzzling and mysterious, appears within our reach. In the past, the utter lack of understanding of these properties of life had been a major factor leading to the view that living organisms possess an inherent, vitalistic "something," perhaps akin to a "spirit." The ability to define this "something" in physical and chemical terms was thought to be unattainable.

One property is the expression and transmission of heritable traits, and the broad outlines of DNA replication and gene action via mRNA synthesis are helping greatly to bring this property to precise definition. Another property is the control of cellular structures; recent experiments support the view that the folding and interactions of structural proteins and many, if not all, of the properties of the fully formed protein are attributable to the primary structure of the molecule. The amino acid sequence certainly limits the number of useful configurations that can be attained in a given intracellular and extracellular environment, and consequently it largely controls the architecture of the living cell.

Gene action and evolution

The mainstream of evolution has been neglected too long in favor of detailed studies on mechanisms of speciation and specialization. The vast majority of extant specialized organisms are the "dead ends" of evolution. They will perish in a changing environment as have more than 99.99 per cent of their specialized ancestors. Thus, evolutionary studies have concentrated on unmasking the forces that lead to "dead ends" and ultimately withdraw organisms from evolution.

What we desire to know more about is the past, the continuity of life. The molecular architecture and mode of functioning of organisms can tell us a great deal about the probable nature of generations that have long since perished. Increasing knowledge of the genetic code, its change in mutation, and its transcription and translation ultimately will assist in accurately interpreting biological events that occurred millions of years ago.

References

Brachet, J., *Chemical Embryology*. New York: Interscience Publishers, 1950.

Brown, D. D., and J. B. Gurdon, "Absence of Ribosomal RNA Synthesis in the Anucleolate Mutant of *Xenopus laevis*," *Proc. Nat. Acad Sci. U.S.*, *51* (1964), 139. An excellent example of how new techniques, utilized in work with microorganisms, may be fruitfully applied to investigation of higher organisms.

Lederberg, J., "Genes and Antibodies," *Science*, *129* (1959), 1649. A provocative article that inquires into the mechanisms underlying antibody formation.

Lewis, E. B., "The Phenomenon of Position Effect," *Advan. Genet.*, *3* (1949), 73. A review of the literature, valuable for its discussion of variegated position effects.

Swanson, Carl P., *Cytology and Cytogenetics*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1957. A thorough, advanced textbook that treats position effects and gene dosage and gives numerous references to the original literature.

Vogel, H. J., and V. Bryson, eds., *Evolving Genes and Proteins.* New York: Academic Press, Inc., 1965. Provocative and contemporary approach to the study of evolution employing the methods of molecular biology.

Wilson, E. B., *The Cell in Development and Heredity,* 3rd ed. with corrections. New York: The Macmillan Company, 1925. This and Brachet's *Chemical Embryology,* classics of the eras in which they were written, can call attention to many interesting problems that current concepts and techniques in molecular biology appear capable of solving in the future.

Index

A

Aberrations (*see* Deletion, Multisite mutation):
  chromosomal, 145
Abortive transduction, 65
Acceptor RNA (*see* sRNA)
Accumulations, 51, 81-82
  cross-feeding, 80, 83
  and feedback inhibition, 87
  inhibitions by, 82-84
Acetylated amino acids, 14, 16
Activating enzyme (*see* Amino acyl RNA synthetase)
Adams, M. H., 4
Adaptor RNA (*see* sRNA)
Adaptive enzyme (*see* Inducible protein)
Adenine, structure, 2, 28
Adenine-thymine base pair, 2, 99
Adenine-uracil base pair, 3
Adenosine triphosphate, structure, 28
Adenylic acid, 28
Alanine, structure, 8
Allele (one form of a gene; *see* Gene)
Allelism (*see* Recombination, Complementation):
  test for, 67
Allosteric effects, 24
  in feedback inhibition, 84-87
  in repression, 102, 103
Alpha helix, 17
Ames, B. N., 75, 106, 111
Amino acid, definition, 7
Amino acids:
  acetylated, 14, 16
  amphoteric nature, 7
  formulas, 8
  general formula, 7
  in protein tertiary structure, 20, 21
Amino acid activating enzyme (*see* Amino acyl RNA synthetase)
Amino acid activation, 27-30
Amino acid incorporation, 38, 40
  and genetic code, 133-35
Amino acid residue, definition, 11
Amino acid sequence:
  methodology, 11
  of TMV protein, 10
Amino acid substitution:
  and genetic code, 124
  relation to function, 60, 61
  relation to phenotype, 114
  in proteins, 56-59
  in TMV protein, 130
  in tryptophan synthetase (*see* Tryptophan synthetase)
Amino acid transfer, 30-32
Amino acid adenylate (*see* Amino acyl adenylate)
Amino acid–sRNA (*see* sRNA)

Amino acyl adenylate:
  reactivity, 29
  structure, 28
  transfer to sRNA, 29, 31
Amino acyl RNA synthetase, specificity, 28, 29
Aminopeptidase, 14
Anfinsen, C. B., 25
Annealling, RNA-DNA, 95
Antibodies (*see* Immune serum)
Arginine, structure, 8
Arginine biosynthesis, 88-89
Asparagine, structure, 8
Aspartic acid, structure, 8
ATP (*see* Adenosine triphosphate)
ATP-pyrophosphate exchange, 29
Auxotroph (*see* Mutant, growth-factor requiring)

B

Bacteria as genetic tools, 49
Bacteriophage:
  mixed infection, 66
  ϕX174, 96
  T4, 114-19
Baker, W. K., 145
Bamford, C. H., 25
Basal enzyme levels, 105
Base pairing, 2, 3, 30-32, 40, 42-43, 94
  in mRNA synthesis, 94
Base sequence (*see* Nearest neighbors in RNA)
Bases, methylated, 30
Beadle, G., 48, 68
Benzer, S., 116, 121, 137
Bifunctional enzymes, 53-55
Biochemical pathways (*see* Pathways)
Bonner, D. M., 72, 111
Brachet, J., 147

C

Carbon disulfide, protection of lysine, 12
Carboxypeptidase, 14
Carlson, E. A., 75
Cells, *frontispiece*
  bacteria, 32
Channeled pathways, 88-89
Chantrenne, H., 44
Charged sRNA, definition, 31
Chemical mutagenesis, 97-100, 135
  triplet coed, 126-27
Chymotrypsin, specificity, 12
Chymotryptic peptides:
  method, 12
  in TMV protein, 13
cis-trans test:
  methods, 64-66
  for operator and regulator genes, 104-105
  relation to polypeptide chains, 66, 67
  with *rII* mutants, 115
  theoretical basis, 67
Cistron, definition, 68
Cistrons, in *rII* region, 115
Chromosome aberrations, 145
Coding ratio, 125-26
Code (*see* Genetic code)
coli (*see Escherichia coli*)
Colinearity of DNA and polypeptide chain, 4, 126, 135-36
Commaless code, 130-31
Competitive inhibition (*see* Inhibition, isosteric)
Complementary genes, 140-41
Complementary RNA (*see* mRNA)
Complementation, 63-75
  and degenerate code, 131
  intragenic, 70-74
    properties of enzyme, 74
    theoretical basis, 73
  in vitro, 73-74
  relation to function, 68-69
  in *rII* region, 115
  theoretical basis, 67
Complementation map, 72, 74
Complementation test, methods, 64-66
Conformation of protein (*see* Protein, conformation)
Conidia, 49
Conjugation, 65
Constitutive mutations:
  in operator, 104-105
  in *R* gene, 107
Constitutive proteins, 102-103
Conti, S. F., 32
Coordinate repression, 105-106
Corepressor, 103
Coupling-repulsion (*see cis-trans* test)
Crick, F. H. C., 128, 129, 137
CRM:
  catalytic abnormalities, 52-53
  definition, 51-52
  and degenerate code, 131
  detection, 52
  in missense and nonsense mutations, 119
  in *Salmonella* mutants, 55
  tryptophan synthetase:
    Escherichia coli, 54
    Neurospora, 51-53
Cross-feeding, 80, 83
Cross-reacting material (*see* CRM)
C-terminal amino acid, definition, 11
Cyanogen bromide, method, 12
Cysteine:
  structure, 8
  disulfide bonds in proteins, 20
Cytosine, structure, 2

D

Davis, R. H., 90
Degenerate code, 113, 125, 131-32
Dehydrase, imidazoleglycerol phosphate, 73-74
Dehydrogenase, L-histidinol, 54-55
Deletion, 72 (*see also* Multisite mutation)
  in operator gene, 107-108
  of repressor gene, 103
  in *rII* region, 116-17
Demerec, M., 145
Deoxyribonuclease, insensitivity of hybrids, 37, 95
Deoxyribonucleic acid (*see* DNA)
Denaturation of proteins, 19
Density gradient centrifugation, 39, 110
Derepression, 101-109
Deuterium exchange, 17
Differentiation, 141-46
Dihybrid ratios, 140-41
Dimer (*see* Protein, quaternary structure)
Dintzis, H. J., 42, 121
Diploid, 140-41
  heterozygous, 64
Diploids, partial, 65
Disulfide bonds, 20
  reduction of, 21
DNA:
  base composition, 1
  heat denaturation, 38, 95
DNA molecule:
  complementary base sequence, 2-3, 96-97
  polarity of chains, 96
DNA-RNA hybrid: 3, 37-38, 94-95
  heat denaturation, 95
  as indication of homology, 95-96, 109
Dominance:
  in *cis-trans* test, 67-68
  of active repressor gene, 103-104
  of operator constitutive, 104-105
Dosage compensation, 142
Dreyer, W. J., 137
Drosophila, 48, 142-45
Duplicate enzymes, 88-89
Duplicated genes, control of function, 107

E

Edman procedure, 15
Electrostatic bonds, 21
Endoplasmic reticulum:
  definition, 33
  electron micrograph, *frontispiece*
End-product deficit, 84
End-product inhibition (*see* Feedback inhibition)
Environment (*see* Inhibition, Repression, Induction)
Enzyme, 6 (*see also* Protein)
  constitutive, 102
  duplicate, 88-89
  inducible, 102
  multifunctional, 53-55
  mutant, properties, 51-55
Enzyme formation:
  repression (*see* Repression)
  basal levels, 105
Ephrussi, B., 48
Escherichia coli:
  electron micrograph, 32
  tryptophan synthetase (*see* Tryptophan synthetase, *Escherichia coli*)
Exopeptidase, 14
Extragenic suppressor mutation, 88-89
Evolution:
  and protein structure, 146
  and gene action, 147
  and degenerate code, 131
Eye-pigment mutant, 48, 145

F

- Feedback inhibition, 24, 84-87
  - competition for substrate, 87-88
  - effect on accumulation, 87
  - histidine pathway, 84-87
  - in vitro, 85-87
  - use of analogues, 85-86
- Feedback-resistant mutant, 85, 86
- Fincham, J. R. S., 4, 75
- "Fingerprint":
  - technique, 13
  - of tryptophan synthetase:
    - A protein of *Escherichia coli*, 14
    - of *Neurospora*, 71
- 5-fluorouracil, 98-100
- Fraenkel-Conrat, H., 22, 58
- Function, 47, 68

G

- Garrod, A., 48
- Genes:
  - active (*see* Induction, Derepression)
  - clustered, 53-56, 105-106
  - complementary, 64-66, 140-41
  - duplicated, 107
  - estimates of size, 63
  - expression (*see* Induction, Repression, Position effect, Dosage compensation)
  - functions, 112
  - inactive (*see* Repression)
  - interaction of (*see cis-trans* test, Complementation, Repression)
  - linked, 53-56, 105-109
  - modifying (*see* Gene, suppressor; Mutation, secondary effects)
  - operational definitions of, 46, 47
  - regulatory, 102
  - for ribosomal RNA, 112-13
  - for sRNA, 112-13
  - spacers between, 115-17, 123-24, 130-31
  - structural, 105-109
- Gene dosage compensation, 142
- Gene-enzyme relationships, 49-62
- Genetic block:
  - accumulations, 81, 82
  - definition, 50
  - growth requirement, 78-80
  - histidine pathway, 78
  - incomplete, 50, 80, 81
  - localization, 77-80
  - partial, 50, 80, 81
  - pathway analysis, 78-80
  - relative position, 78, 82
  - tryptophan pathway, 78
- Genetic code, 123-27
  - amino acid incorporation studies, 133-35
  - amino acid substitutions, 124
  - coding ratio, 125-26
  - commaless, 130
  - deciphering, 124-25
  - degenerate, 113, 125, 131-32
  - intragenic suppression, 128-29
  - missense, 113
  - modulating triplets, 120-21
  - nonoverlapping, 130
  - nonsense, 113
  - and recombination, 135-36
  - synthetic ribopolynucleotides, 124-25, 133-35
  - terminology, 113-14
  - triplet, 125-26, 133-35
  - triplet composition, 135-36
  - triplet sequence, 135
  - tryptophan synthetase, 135-36
  - types, 123-24
  - universality, 132-33
- Genetic homology, indication of, 95-96
- Genetic map (*see* Map, genetic)
- Genotype, 59
  - definition, 6
- Glutamic acid, structure, 8
- Glutamine, structure, 8
- Glycine, structure, 8
- Gorini, L., 121
- Growth of mutants, 49, 78-79
- Growth factors for mutants, 49, 78-80
- Guanine, structure, 2
- Guanine-cytosine base pair, 2, 3
- Guanosine triphosphate in protein synthesis, 38

H

- Hayes, W., 4
- Hall, C. E., 33, 35
- Helix:
  - alpha, 17
  - in DNA, 2
  - in sRNA, 30-31
- Hemoglobin:
  - biosynthesis, 41, 42
  - effects of mutation, 90
  - tertiary structure, 18, 19
  - quaternary structure, 19
- Heterokaryon, 65, 72
- Heterochromatin, 143-44
- Heterogenote, 65
- Heterozygote, 64-66, 140-41, 142
- Histidine, structure, 8, 79
- Histidine biosynthesis, 54-56, 79
  - enzymes, 57, 79
  - intermediates, 57, 79
- Histidine mutants, complementation tests, 69
- Histidine region:
  - control, 105-109
  - genetic structure, 57
  - modulation, 120
- Histidinol dehydrogenase, 54-55
- Histidinol phosphophosphate phosphatase, 55, 73-74
- Histones, 142
- Horne, R. W., 22
- Hybrid complex, DNA-RNA (*see* DNA-RNA hybrids)
- Hyperrepressed, 103
- Hyphae, 49
- Hydrogen bonds:
  - in alpha helix, 17
  - between base pairs, 2
  - between protein and water, 17
  - in tertiary structure, 20-21
- Hydrophobic interactions, 21
- Hydroxyproline:
  - biosynthesis, 16
  - structure, 8

I

- Inducer, 103
- Inducible protein, 102
- Induction, 101-109
- Ingram, V. M., 61
- Inhibition:
  - allosteric, 24, 84-87
  - competitive (*see* Inhibition, isosteric)
  - effect of accumulations, 82-84
  - feedback, 24, 84-87
  - isosteric, 24, 88
- Intercalation, 126-27
- Intragenic complementation, 70-74
- Immune serum, preparation, 52
- Ionic bonds, 21
- Isoleucine, structure, 8
- Isomer, optical activity, 7
- Isosteric effect, 24, 88

J

- Jacob, F., 5, 111

K

- Kendrew, J. C., 18, 25

L

- Leaky mutant, 50, 80, 81
- Lederberg, J., 147
- Lerman, L. S., 127
- Leucine, structure, 8
- Lewis, E. B., 147
- Lewis effect (*see cis-trans* test)
- Linkage, 46
- Linked genes, 105-109
- Linkage map (*see* Map, genetic)
- Locus (position occupied by gene; *see* Gene)
- Lysine, structure, 8

M

- Map:
  - complementation, 72, 74
  - genetic:
    - *Drosophila*, 145
    - *Escherichia coli* tryptophan synthetase, 4, 56
    - histidine region, 57
    - *Neurospora* tryptophan synthetase, 72
- Martin, R. G., 110, 111
- Mendel, G., 47
- Messenger RNA (*see* mRNA)
- Methionine:
  - cleavage at, 12
  - structure, 8
- Methylated bases, in sRNA, 30
- Microsomes, 33
- Missense, 113, 117-18
- Mitchell, H. K., 5
- Mixed infection, 66
- Modifier (*see* Mutation, suppressor; Mutation, secondary effects)
- Modulation, 119-21
- Monod, J., 111

Mosaic:
  due to mRNA mutation, 97-98
  variegated position effect, 143-45
mRNA, 36-38, 40
  attachment to ribosomes, 39, 40
  biosynthesis, 36
  detection, 39, 109
  in development, 141
  general regulation of synthesis, 100
  for histidine operon, 109-10
  induced mutations, 97-98
  in induction, 109
  multicistronic, 109, 119-20
  multigenic, 109, 119-20
  mutations, 97
  properties, 36
  pulse labeling, 39, 110
  quantitation, 109
  regulation of synthesis, 93-111
  relation to operon, 109-10
  single-strandedness, 96
  size, 109-10
  stability, 36
  synthesis and degradation, 36, 94, 96-97, 109
  synthetic:
    and degenerate code, 131
    limitations, 134-35
  as template, 40
  in vitro synthesis, 36
Mutagenesis, chemical, 97-100, 126-27, 135
Multifunctional protein, 53-55
Multigenic (*see* Operon; mRNA, multigenic)
Multisite mutation, 69
Mutant:
  CRM-negative, definition, 51-52
  CRM-positive, definition, 51-52
  enzyme deficiency, 51
  eye-pigment, 48, 145
  feedback-resistant, 85-86
  growth-factor requiring, 49
  leaky, 50, 80-81
  operator constitutive, 104
  operator negative, 104
  rII, 114-19, 127-29
  selection of, 49
  subvital (*see* Mutant, leaky)
  temperature-sensitive, 51
  variegated, 145
Mutant protein, 46-62
  properties, 51-55
Mutation, 47
  induced, 97-100, 126-27, 135
  missense, 113
  in mRNA, 97-98
  multiple effects of, 6
  multisite, 69
  nonsense, 113, 118-19
  operator constitutive, 107
  polarity, 120
  reading-frame, 128-29, 132
  in repressor gene, 103
  secondary effects, 77-92
  suppressor:
    extragenic, 88-89
    intragenic, 60, 117-18, 128-29
  transition, 135
  transversion, 135

N

N-bromosuccinimide, 12
Nearest neighbors in RNA, 37
Neurath, H., 25
Neurospora:
  as genetic tool, 48
  growth requirements, 49
  heterokaryon, 65
  terminology, 49
  tryptophan synthetase (*see* Tryptophan synthetase, *Neurospora*)
Nirenberg, M. W., 134, 137
Noncovalent bonds, types in protein structure, 17
Nonoverlapping code, 123-24, 130
Nonpolar side chains, 21
Nonsense, 113, 118-19
nRNA (*see* mRNA)
N-terminal amino acid:
  definition, 11
  acetylated, 14, 16
Nucleic acids (*see* DNA, mRNA, sRNA, Ribosomal RNA)
Nucleolus, *frontispiece*, 142
Nucleoside, structure, 28
Nucleoside triphosphate, structure, 28
Nucleotide, structure, 28
Nucleus, *frontispiece*
Nutritional requirements:
  of *Neurospora*, 49
  of mutants, 49, 78

O

One gene, one enzyme, 48
One gene, one polypeptide chain, 47-48, 60, 69-70
Operator, 102-10
  affinity for repressor, 104-105
  in *cis-trans* test, 104-105
  deletion, 107-108
  function, 104-105
  mutation, 104
  relation to structural gene, 105-109
Operator constitutive mutation, 104, 107
Operator negative mutant, 104
Operator site, 103
Operator zero mutant, 104
Operon, 105-109
  definition, 105
  histidine biosynthesis, 105-109
  relation to mRNA, 109-10
Overlapping code, 124

P

Pathways:
  analysis, 78-80, 82
  channeled, 88-89
  inhibitions by accumulations, 83-84
Pauling, L., 16, 17
Penicillin method, 49
Peptides (*see also* Tryptic peptides):
  sequence of amino acids, 14
  placing in sequence, 16
  separation, 13
Pepsin, specificity, 12, 13
Peptide bond:
  definition, 9
  biosynthesis, 38
Peptide polymerase, 38, 43
Periods between genes, 115-17, 123-24, 130-31
Perutz, M. F., 18, 19, 25
Peters, J. A., 5
Phage (*see* Bacteriophage)
Phenotype, 6, 59
Phenylalanine, structure, 8
Phenylisothiocyanate procedure, 15
Phenylketonurea, 83-84
Phosphatase, L-histidinol phosphate, 55, 73-74
Pigment (*see* Eye-pigment mutant)
Pleiotropy (*see* Secondary consequences of mutations)
Polarity in translation, 120
Polymer (*see* Polypeptide chain; Protein, subunits; Ribopolynucleotide)
Polynucleotide (*see* Ribopolynucleotide)
Polynucleotide phosphorylase, 124-25
Polypeptide chain:
  attachment to sRNA, 41
  definition, 9
  general formula, 9
  growth of, 41-42
  ribosomal-bound, 41, 42
  structure, 17
  synthesis, 41-42
  terminology, 11
Polyphenylalanine, 130
Polysomes, 34-35, 39-40
  in modulation, 120
Position effect;
  cis-trans, 67
  variegated, 143-45
Precursors, 51
  accumulations, 81, 82
Primary structure of proteins, 7-16
  mutational alterations, 55-59, 135-36
Primer for RNA polymerase, 36, 38
Proflavin, 126-27
Proline:
  structure, 8
  action of aminopeptidase, 14
  role in protein secondary structure, 19
Protein:
  abnormal, 46-62
  altered, 51-52
  amino acid content, 9
  conformation, 18-25
    allosteric effects, 24, 84-87, 102-103
    and repression, 102-103
  constitutive, 102, 103
  control of formation (*see* Repression)
  denaturation, 19
  inducible, 102
  modifications in primary structure, 16
  molecular weights, 9
  mutant, 46-62
    properties, 51-55

  polymers, 21-24, 74-75
  primary structure, 7-16
    alterations in, 55-59
  quaternary structure, 21-24, 71-75
  repressor, 102-105
  secondary structure, 16-18
  structure, 6-26
    and evolution, 146
  subunits, 21-24, 74-75
    identical, 71, 72
    importance, 74-75
    non-identical, 70-71
    relation to complementation, 66-75
    tryptophan synthetase, 54, 69-70
  tertiary structure, 18, 21
Protein biosynthesis (*see* Protein synthesis)
Protein synthesis, 27-45
  general regulation, 100
  importance of base pairing, 3, 30-32, 40, 42-43, 94
  over-all scheme, 42-43
Protein-forming system, 38-44
Protein-protein interactions, 22, 74-75
  as basis for complementation, 73
  tryptophan synthetase, 69-70
Proteolytic enzymes, 12, 14
Pseudo-wild, 65
Pulse-labeling of mRNA, 39, 110
Punnett, R. C., 140
Pyrimidine biosynthesis, 88-89
Pyrimidines (*see* Thymine, Uracil, Cytosine)

R

R gene, 102-104
  dominance, 103, 104
Reading frame mutation, 128-29, 132
Recombination, 46-47
  and genetic code, 4, 135-36
  maps, 4, 56, 57, 72
Recessiveness, concept of, 67-68
Regulation on ribosome, 112-22
Regulator gene (*see R* gene, Repression)
Regulatory systems, 102-109
Repression, 101-109
  coordinate, 105-106
  noncoordinate, 106-107
  of unlinked genes, 106-107
Repressor, identity, 102-105
Repressor gene, 102-104
  mutations, 103
Repressor molecules, conformation, 102, 103
Repulsion (*trans* arrangement; *see cis-trans* test)
Reticulocytes, 41, 42
Reversion, second-site, 117-18
Ribonuclease:
  insensitivity of hybrids, 37, 95
  and sRNA structure, 30
Ribonucleic acid (*see* mRNA, sRNA, Ribosomal RNA, Ribopolynucleotide)
Ribopolynucleotide, synthetic, 40, 130, 133-35
Ribosomal RNA:
  genetic control of, 33, 112-13
  properties, 33
Ribosome, 32-35
  alteration in structure, 118
  attachment of sRNA, 34-35, 41
  composition, 32
  electron micrographs, 32, 33, 34
  properties, 34
  regulation on, 112-22
  role in protein synthesis, 34-45
Rich, A., 35
RNA (*see* sRNA, mRNA, Ribosomal RNA)
RNA-polymerase, 36, 95
  comparison to polynucleotide phosphorylase, 125
  and repression, 102, 104-105
rII mutants, 114-19, 127-29

S

Salivary gland chromosome, 144-45
Salmonella histidine enzymes (*see* Histidine biosynthesis)
Sanger, F., 11
Schweet, R., 45
Secondary consequences of mutations, 77-92
Secondary protein structure, 16-18
Second-site reversion, 117-18
Sequencing peptides, 15
Serine, structure, 8
Sex chromosomes, 142
Soluble RNA (*see* sRNA)
Spacers between genes, 115-17, 123-24, 130-31

sRNA, 3
  amino acid acceptor site, 30, 31
  as adaptor, 30-32, 38-41
  attachment to polypeptide, 41
  attachment to ribosomes, 34-35, 41
  base content, 30
  base pairing with mRNA, 32
  and degenerate code, 31, 131
  general structure, 30
  genetic control of, 30, 31, 112-13
  heterogeneity, 31
  terminal trinucleotide, 30, 31
Stahman, M. A., 25
Stent, G. S., 5
Stocker, B. A. D., 98
Structural gene, 105-109
Subtilisin, 12, 13
Substitutions, amino acid, 56-59
Subunits, protein (*see* Protein, subunits)
Sucrose density gradient centrifugation, 39, 110
Sulfhydryl group, 20
Suppressor mutation:
  extragenic, 88-89, 118-19
  intragenic, 60, 117-18, 128-29
Swanson, C. P., 147
Synthetic ribopolynucleotides, 40, 124-25, 131
Syntrophy, 80, 83

T

Tatum, E., 48
T4 bacteriophage, 114-19
td gene (*see* Tryptophan synthetase, *Neurospora*)
Temperature-sensitive mutant, 51
Template:
  for mRNA synthesis, 3, 96-97
  mRNA in protein synthesis, 40
Tertiary protein structure, 18-21
Threonine:
  structure, 8
  C-terminal in TMV protein, 14, 23
Thymine, structure, 2
Tissieres, A., 38
TMV (tobacco mosaic virus):
  amino acid substitutions, 58, 59, 130
  infected leaf, 58
  N-terminal amino acid, 14
  particles, 22, 23
  primary structure of protein, 10
  protein subunits, 22-23
  RNA, 23
  tryptic peptides, 13
trans configuration (*see cis-trans* test)
Transcription, 3, 93-111, 123
Transduction, 65
Transfer RNA (*see* sRNA)
Transition mutation, 99, 135
Translation, 3, 112, 123
Translocation, 145
Transversion mutation, 135
Triplet code, 125-26, 133-35
  chemical mutagenesis, 126-27
  genetic experiments, 127-29
Trypsin, specificity, 12
Tryptic peptides:
  definition, 12
  method, 12
  placing in sequence, 15-16
  polypeptide chain growth, 41-42
  separation, 12
  in TMV protein, 13
Tryptophan:
  biosynthetic pathway, 49-50
  cleavage at, 12
  structure, 8
Tryptophan synthetase:
  Escherichia coli:
    amino acid substitutions, 56-58
    colinearity of gene and protein, 4
    complementation tests, 69-70
    CRM, 55-58
    double mutants, 57
    "fingerprint," 14
    and genetic code, 135-36
    mutant proteins, 54-56
    A protein gene, 55
    proteins of revertants, 59, 60
  Neurospora:
    complementation, 71-72
    CRM, 52-53
    gene, 49
    mutant proteins, 52-53
    position in pathway, 49, 50
    structure of, 71
  reactions catalyzed, 53
Tyrosine, structure, 8

U

Umbarger, H. E., 91

V

Valine, structure, 8
Variegated position effect, 143-45
Vitalism, 146
Vogel, H. J., 111, 137, 147

W

Wagner, R. P., 5
Watson, J. D., 38
Watson-Crick base pairing, 2
  in mRNA synthesis, 94
Weiss, L., *frontispiece*
Wilson, E. B., 147
Wittmann, H. G., 58

X

X chromosome, 142
X-ray diffraction, 18, 23

Y

Yanofsky, C., 4, 56, 61, 121
Y chromosome, 142

Z

Zamecnik, P. C., 45
Zubay, G., 30

Philip E. Hartman and Sigmund R. Suskind

Philip E. Hartman (Ph.D., University of Pennsylvania) and Sigmund R. Suskind (Ph.D., Yale University) are Associate Professors of Biology at The Johns Hopkins University, where Dr. Suskind is also a staff member of the McCollum-Pratt Institute. Both have published numerous studies in the areas of microbial and biochemical genetics.

Prentice-Hall
FOUNDATIONS OF MODERN *Genetics* SERIES
Edited by
Sigmund R. Suskind and Philip E. Hartman

AGRICULTURAL GENETICS
James L. Brewbaker

GENE ACTION
Philip E. Hartman
Sigmund R. Suskind

EXTRACHROMOSOMAL INHERITANCE
John L. Jinks

DEVELOPMENTAL GENETICS*
Clement Markert

HUMAN GENETICS
Victor A. McKusick

POPULATION GENETICS AND EVOLUTION
Lawrence E. Mettler

THE MECHANICS OF INHERITANCE
Franklin W. Stahl

CYTOGENETICS
Carl P. Swanson
Timothy Merz
William J. Young

*Published jointly in Prentice-Hall's *Foundations of Developmental Biology Series*